



UGANDA CHRISTIAN
UNIVERSITY

A Centre of Excellence in the Heart of Africa

Department of Computing & Technology

*Interpretable Predictive Modeling of Postpartum Surgical Site
Infections*

By

Emmanuel Nkurunziza
BSc.SE, MUK

Reg No: *M24M19/007*, Student No: *B28744*

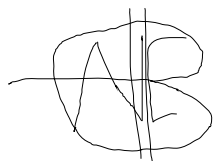
E-mail: *enkurunziza97@gmail.com*, Phone No: *+256776548083*

A dissertation submitted to the Faculty of Engineering Design &
Technology for the study leading to the Degree in partial fulfillment
of the requirements for the award of **Master of Science in Data
Science** of Uganda Christian University

March 2026

Declaration

I, *Emmanuel Nkurunziza*, declare to the best of my knowledge that the work in this dissertation is original and has not been submitted before to any university/institution. It is the result of my effort.

A handwritten signature in black ink, appearing to be 'EN', written over a horizontal line.

Signature: _____

Date: 11/06/2026

Ethical Generative AI Usage Declaration

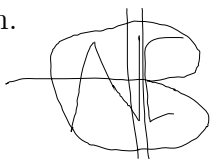
Students may use generative artificial intelligence (AI) ethically as a “study buddy” or assistant rather than a replacement for critical thinking. This includes citing AI use and fact-checking all outputs to maintain academic integrity.

For purposes of this assignment, the following shall guidelines shall apply.

- i. Generative AI may be used for brainstorming, and generating ideas to improve your work. You may also use AI for literature search, and summarizing journal papers for quick analysis and assessing relevance of publications. However, once the relevant publications have been identified, a student is expected to fully review the selected publications, identifying problems addressed, methods used, gaps identified and recommendations made.
- ii. Where AI has been used as guided in (i) above, full disclosure of how it has been used should be made in the appendix C, including how AI was used, the AI tools (including mode versions), and a compilation on major AI prompts used.
- iii. NO GENERATIVE AI CONTENT IS ALLOWED IN THE FINAL SUBMISSION. The final submission SHOULD be your original work.

A. Declaring the non-use of AI for content generation

I, *Emmanuel Nkurunziza*, declare that no content generated by AI technologies has been used in this dissertation.



Signature: _____ Date: 11/06/2026

B. Acknowledging the use of AI as a study buddy

Please refer to Appendix C for the complete declaration of AI usage for generating ideas, brainstorming, literature search, and summarizing only to improve your work.

Approval

This dissertation has been submitted for examination with approval of the following advisors/supervisor.

Advisor: *Dr. Terhemba Michael-Ahile*
Department of Computing & Technology
Faculty of Engineering Design & Technology
Uganda Christian University

Signature:  _____

Date: **11/06/2026** _____

Abstract

Background:

Postpartum surgical site infections (SSIs) are a major cause of maternal morbidity in low-resource settings, yet early detection is limited by reactive, symptom-based surveillance systems. Machine learning (ML) and explainable artificial intelligence (XAI) offer potential tools for proactive SSI risk identification using routinely collected clinical data.

Methods:

A retrospective dataset of 8,918 caesarean deliveries from three regional referral hospitals in Uganda (SSI prevalence: 8.1%) was analysed. After rigorous cleaning, leakage-free feature engineering, imputation, and encoding, four ML models Logistic Regression, Random Forest, Support Vector Machine (RBF), and XGBoost were trained and evaluated under extreme class imbalance using SMOTENC. Model performance was assessed using accuracy, precision, recall, F1-score, ROC-AUC, and precision–recall curves. SHAP explainability was applied to identify globally and locally influential predictors.

Results:

Overall discriminatory performance was modest due to limited predictive signal in routine clinical records. Logistic Regression achieved the highest sensitivity (recall = 0.9931; ROC-AUC = 0.6528), whereas XGBoost produced the highest accuracy (0.7932) but poor recall (0.2639). SHAP analysis highlighted preoperative showering, skin antiseptic preparation, obstructed labour, and prolonged surgery duration as key contributors to increased SSI risk.

Conclusion:

Explainable ML models are feasible for early SSI risk identification in low-resource settings but are constrained by sparse routine data. Logistic Regression may serve as a high-sensitivity clinical screening tool. Improved performance will require enriched clinical datasets, inclusion of intraoperative variables, and multi-centre validation to support practical deployment in maternal health surveillance.

Keywords:

Surgical Site Infection, Machine Learning, Explainable AI, SHAP, SMOTENC, Maternal Health, Uganda.

To my family, whose sacrifices shaped my path and whose belief in me kept me going. To my academic advisor, who guided me from the start to the end of this project, nurtured my curiosity and challenged me to grow. And to every young scholar who dares to dream may this work remind you that persistence creates possibilities.

Acknowledgments

First, I would like to thank the almighty God for enabling me finish this master's degree timely. Secondly, I promised my wife that if I ever completed this master's degree, I would mention her first, above all others, to make it clear whose support mattered most. To my wife Immaculate, thank you for your unwavering love, encouragement, and the smile that kept me grounded and motivated throughout this journey.

I am deeply grateful to my academic advisor, Dr. Terhemba Michael Ahile, for her guidance, mentorship, and the many opportunities she provided that helped me grow in confidence as a data scientist. Thank you for believing in this study and for generously sharing your wisdom. My work is undoubtedly stronger because of you.

My sincere appreciation goes to the Active Prevention and Treatment (APT) of Maternal SEPSIS team whose support made this master's program possible. Thank you, Professor Catriona Waitt, David Lissauer, Peter Waitt, and Peace Okwaro, for providing the studentship scholarship and for your commitment to capacity building.

I also extend heartfelt gratitude to the Healthcare Associated Infections (HAI) team and the leadership of the Infectious Diseases Institute, my professional home for the past eleven years. Thank you for allowing me to use the organizational data for this study and for shaping me into the professional I am today.

To my children Garvin, Elijah, Elisha, and Elsie thank you for brightening every day and for joyfully interrupting my attempts to concentrate. You are my greatest source of happiness, and Daddy loves you deeply. Finally, to my wife Immaculate once more thank you. You truly are the best, and I am saving the best for last.

Contents

Declaration	i
Ethical Generative AI Usage Declaration	ii
Approval	iii
Abstract	iv
Dedication	v
Acknowledgments	vi
1 Introduction	1
1.1 Background	1
1.2 Problem Statement	2
1.3 Research Objectives	2
1.3.1 General Objective	2
1.3.2 Specific Objective	3
1.4 Research Questions	3
1.5 Justification	3
2 Literature Review	4
2.1 Introduction	4
2.2 Clinical Background of Postpartum Surgical Site Infections (SSIs) in LMICs . .	4
2.3 Risk Factors and Early Detection Challenges	5
2.3.1 Well-Known Clinical and Procedural Risk Factors	5
2.3.2 Patient-Related Risk Factors	5
2.3.3 Postoperative Factors	5
2.3.4 Challenges in Early Detection of SSI	6
2.3.5 Rationale for Machine Learning in SSI Detection	6
2.4 Machine Learning for Surgical Site Infections	6
2.4.1 Predictive Models and Performance	6
2.4.2 Data Utilization	7
2.4.3 Implications for Healthcare	7
2.5 Explainable Artificial Intelligence (XAI)	8
2.6 Machine Learning and Explainability in Low resource setting	8
2.7 Explainable Artificial Intelligence (XAI) in Medical Care	9
2.8 Explainability and ML Based Predictive Modelling	10
2.8.1 Research priorities and open problems	10
2.9 Research Gaps Identified	10

2.10	Conceptual Framework	13
3	Methodology	14
3.1	Introduction	14
3.2	Study Design	14
3.3	Data Source	14
3.3.1	Data Collection and Acquisition	14
3.3.2	Dataset Description	15
3.3.3	Data Governance and Inclusion Criteria	15
3.4	Data Management Plan	16
3.5	Data Cleaning and Preprocessing	16
3.5.1	Missing Data Analysis	17
3.5.2	Duplicate data	21
3.6	Exploratory Data Analysis (EDA)	21
3.6.1	Univariate Analysis	21
3.6.2	Correlation Matrix	25
3.6.3	Univariate analysis of Categorical Variables	25
3.6.3.1	Distribution of C-Section Type	27
3.6.3.2	Distribution of Indications for Caesarean Section	28
3.6.3.3	Distribution of Caesarean Sections by Surgical Personnel	30
3.6.3.4	Distribution of Caesarean Sections by Anesthesia Personnel	31
3.6.3.5	Distribution of HIV Status among Patients	32
3.6.3.6	Distribution of Patient Skin Preparation Methods	33
3.6.4	Data Structure	33
3.7	Analytical Environment	34
3.7.1	Overview of Analytical Pipeline	34
3.8	Feature Engineering	36
3.9	Model Development	38
3.10	Model Performance Evaluation and Validation	38
3.11	Model Interpretability and Insights	39
3.12	Alignment Between Objectives and Methods	39
3.13	Validation and Clinical Contextualization	39
3.14	Ethical Considerations	40
4	Results	41
4.1	Descriptive statistics	41
4.1.1	Bivariate Analysis of the numerical variables with the Target Variable	42
4.1.2	Examining the Target Variable	42
4.2	Model design and Performance comparison	43
4.2.1	Model Development	43
4.2.2	Model Performance Evaluation	44
4.2.3	Model Performance Results	45
4.2.4	ROC Curves for all the models	45
4.2.5	Precision –Recall Curve Analysis for all the models	46
4.2.6	Confusion Matrix Analysis	48
4.2.7	Feature Importance Results	49
4.3	Explainability using Shapley Additive exPlanations (SHAP)	50
4.3.1	SHAP bee swarm plot summary	50
4.3.2	SHAP Waterfall Plot for Individual SSI Risk Explanation	52
5	Discussion	54

5.1	Interpretation of Key Findings	54
5.2	Comparison with Literature	54
5.3	Implication for Practice - Clinical Practice, Policy, Machine Learning Modelling.	55
5.4	Methodological Reflections	55
5.5	Study Limitation	55
6	Conclusion and Recommendation	56
6.1	Recommendation	56
6.1.1	Enhance Data Collection and Feature Scope	56
6.1.2	Refine Modelling Approaches for Imbalanced Data	56
6.2	Future Work	56
6.3	Conclusion	57
	REFERENCES	61
	APPENDICES	61
A	Budget	62
B	Research Project timelines	63
C	Declaration of AI use	64
D	List of 10 Major Journal Publications reviewed	65
E	Research alignment with national (NDPIV) and Global (SDGs) agenda	66
E.1	Alignment with the National Development Plan IV (NDPIV)	66
E.1.1	Alignment with NDPIV	66
E.1.2	Alignment with the Uganda Vision 2040	66
E.1.3	... other agendas	66
E.2	Alignment with global agenda	66
E.2.1	Alignment with Sustainable Development Goals (SDG)	66
E.2.2	Alignment with	66
E.2.3	Alignment with	66
F	Algorithms, Code of interest	67
G	Research poster layout	68
H	Research Ethics Committee (REC) Approval	69
I	Data collection tools	70
J	Student - supervisor meeting/supervision plan	71

List of Figures

3.1	Number of missing values per feature in the dataset.	17
3.2	Missing data matrix illustrating the pattern of missingness across all variables in the dataset.	19
3.3	Correlation heatmap showing the pairwise relationships among variables with missing data in the surgical site infection (SSI) dataset.	20
3.4	Hierarchical clustering dendrogram illustrating similarity relationships among variables in the surgical site infection (SSI) dataset.	20
3.5	Distribution plots for key numerical variables in the dataset.	22
3.6	Boxplots of selected numerical variables.	23
3.7	Updated distribution plots for key numerical variables after handling outliers.	24
3.8	Updated boxplots for key numerical variables after handling outliers.	24
3.9	Correlation heatmap illustrating pairwise relationships among the cleaned numerical variables	25
3.10	Distribution of surgical cases across study sites.	26
3.11	Distribution of ASA scores among patients.	26
3.12	Distribution of patients across Age-groups.	27
3.13	Distribution of C-section types showing the number of emergent and elective cases.	28
3.14	Distribution of indications for C-section showing patient counts across multiple clinical causes.	29
3.15	Distribution of surgical personnel involved in C-sections, showing the number of patients managed by each cadre.	30
3.16	Distribution of cadres of surgical personnel involved in Caesarean sections.	31
3.17	Distribution of HIV Status among Patients	32
3.18	Distribution of Patient Skin Preparation Methods	33
3.19	Overview of the analytical pipeline for Surgical Site Infection (SSI) prediction.	36
4.1	Boxplots illustrating the distribution of key continuous variables stratified by surgical site infection (SSI) outcome.	42
4.2	Distribution of Surgical Site Infection (SSI) Diagnosis Among Patients	43
4.3	ROC Curves comparing Logistic Regression, Random Forest, SVM (RBF kernel), and XGBoost models.	45
4.4	Precision–Recall curves for Logistic Regression, Random Forest, SVM (RBF kernel), and XGBoost models.	46
4.5	Confusion matrices illustrating the classification performance of Logistic Regression, Random Forest, Support Vector Machine (RBF kernel), and XGBoost models.	48
4.6	Top 20 most influential features for predicting Surgical Site Infections (SSI).	50
4.7	SHAP Summary Plot Showing Feature Impact on Model Output	52
4.8	SHAP Decision Plot Illustrating Feature Contributions Toward the Predicted Output.	53

A.1	Budget allocation for the study activities.	62
B.1	Activity Timeline	63

List of Tables

2.1	Research Gap Summary	12
3.1	Summary of Dataset Characteristics	15
3.2	Summary of Missing Values per Feature	18
3.3	Software Tools and Libraries Used in the Study	34
3.4	Summary of Objectives, Methods, and Outputs	39
4.1	Performance Evaluation of the Models	43

Chapter 1

Introduction

1.1 Background

Surgical site infections (SSIs) remain among the most frequently reported healthcare-associated infections globally and impose a disproportionate burden on low- and middle-income countries (LMICs) [1]. Their occurrence is more than three times higher than in high-income countries, leading to prolonged hospital stays, frequent readmissions, and increased healthcare expenses for both patients and the healthcare system as a whole [1, 2].

Among women undergoing Caesarean section, surgical site infections are one of the leading causes of postpartum morbidity and can contribute to delayed recovery, prolonged antibiotic use, and increased risk of maternal complications, with a global incidence ranging from 3% to 15% [3]. Studies further show that a considerable proportion of postpartum SSIs develop after hospital discharge and are frequently detected during follow-up visits rather than during inpatient care [4].

In Uganda and other low-resource settings, healthcare-associated infection surveillance systems remain underdeveloped, and routine post-discharge follow-up is inconsistently implemented [5]. Consequently, the true burden of postpartum SSIs may be underestimated. Existing surveillance reports from Hoima, Kayunga, and Entebbe Regional Referral Hospitals indicate increasing detection of SSI cases during day 6 and day 14 follow-up visits after discharge, suggesting that infections are frequently identified only after clinical symptoms have progressed. However, many SSIs are preventable through infection prevention strategies and good surgical practices; the increased healthcare costs and associated patient morbidity from SSIs are avoidable in most settings.

A properly functioning surveillance and control system can help healthcare providers to reduce about 10% to 35% of surgical site infections [5, 6]. Therefore, through the Infectious Diseases Institute and funding of the US CDC, the Ministry of Health, Uganda, is implementing a national surveillance program pilot for healthcare-associated infections (HAIs) as outlined in WHO Core Component 4, with the SSI among women post-partum as the main focus. This surveillance is to provide valuable information on ways to inform the implementation and ongoing evaluation of SSI prevention interventions, improve patient outcomes, and ensure patient safety in the long run.

Detecting and treating SSI early with a clear established criterion has been recommended in several studies as a way to reduce the burden of healthcare-associated infections, to inform patient care and proper sepsis treatment, and management among women postpartum or post-abortion [7, 8]. The application of machine learning (ML) techniques in predictive analytics

using data from available health systems to predict surgical site infections has received attention for early intervention[9]. Several machine learning algorithms have been applied to improve early diagnosis of SSIs among women post-partum.

However, one of the main challenges with ML is the “black-box” nature of many models, which limits their acceptance in clinical practice[10]. To address this, explainable AI (XAI) methods such as SHAP and LIME are increasingly being used to make model predictions transparent and interpretable, helping clinicians understand which clinical and demographic factors contribute most to infection risk. This not only improves trust in predictive tools but also supports better decision-making and integration into routine surveillance systems[11].

The HAI CDC pilot Study is a multi-site research initiative aimed at establishing an SSI surveillance program in three regional referral hospitals (RRH) in Uganda. This research will make use of data collected from these sites to build and compare multiple machine learning models to identify or predict SSI occurrence early, before day six (6) of the recommended review day for C-section mothers’ post-partum. It will also assess the performance of these models to determine the most effective one for enhancing maternal health outcomes in low-resource environments.

1.2 Problem Statement

Postpartum surgical site infections (SSIs) detected on day 6 and day 14 after discharge have increased in the surveillance reports from Hoima, Kayunga, and Entebbe Regional Referral Hospitals, revealing a critical gap in early detection. Current surveillance approaches are largely reactive, identifying infections only after symptoms emerge, which delays timely intervention and increases maternal morbidity.

Early identification of SSIs is essential for improving clinical decision making and strengthening national infection prevention systems. Predictive tools, including machine learning methods combined with explainable artificial intelligence, offer the potential to support proactive surveillance by highlighting patients at elevated risk and clarifying the factors driving these risks.

However, existing approaches do not reliably predict which postpartum women are most vulnerable, and clinicians require transparent, interpretable outputs to trust machine learning predictions. This study therefore seeks to evaluate the performance of explainable machine learning models for early SSI prediction and identify the clinical and demographic features most associated with infection risk, with the goal of informing more effective surveillance and targeted prevention strategies. .

1.3 Research Objectives

1.3.1 General Objective

To assess the feasibility of explainable machine learning models for early detection of postpartum Surgical Site Infections(SSI), and to identify data gaps in routine hospital records that limit predictive performance.

1.3.2 Specific Objective

1. To identify and evaluate the predictive value of routine clinical and demographic features for postpartum SSI risk.
2. To compare the performance of different ML models given the current hospital data constraints.
3. To apply explainability techniques to highlight feature importance and clarify data gaps for clinicians.

1.4 Research Questions

1. Which patient clinical and demographic features contribute most to the predictions of machine learning models for postpartum SSIs?
2. Which machine learning model performs best in the early detection of postpartum surgical site infections (SSIs) among mothers in Uganda?
3. How can model explainability techniques help interpret machine learning predictions for postpartum SSIs in Ugandan healthcare settings?

1.5 Justification

SSI rates in LMICs range from 8 to 30% of procedures, making them the most frequent Healthcare Associated Infections (HAIs) with substantial morbidity, mortality, and economic impacts on mothers after C-section. Caesarean deliveries have been steadily increasing in most industrialized countries and many emerging countries, including Uganda, over the previous few decades, which has raised concerns among social scientists[3].

In recent years, it has been suggested that obstetricians are increasingly favoring surgical birth over natural birth. In addition, there is some indication from specific nations that mothers who wish to deliver their child via C-section are becoming more adamant about their preferences[2]. Conducting surveillance for SSIs is recognised as making an important contribution to reducing the risk of these infections[3] and establishing high-quality surveillance with timely feedback to surgeons can lead to a reduction in the risk of SSI.

This surveillance will allow for the early detection and prompt management of postoperative infections, reducing the risk of complications, such as sepsis, wound dehiscence, and endometritis. By identifying and addressing these infections promptly, surveillance will help save lives and improve health outcomes for both mothers and infants.

Chapter 2

Literature Review

2.1 Introduction

Machine learning is increasingly being applied to healthcare problems, including infectious disease prediction and maternal health outcomes[12]. Postpartum surgical site infections (SSIs) remain a major cause of maternal morbidity and mortality, particularly in low- and middle-income countries such as Uganda. This chapter reviews existing literature on SSIs, the role of early detection, applications of machine learning in healthcare, explainability in AI, and the contextual challenges in Ugandan healthcare settings.

2.2 Clinical Background of Postpartum Surgical Site Infections (SSIs) in LMICs

Surgical site infections (SSIs) are defined as infections that occur within 30 days after a surgical procedure, including after Caesarean sections, which are among the most common obstetric surgeries worldwide[13]. Globally, SSI rate after a surgical procedure ranges from 3% to 15% worldwide, with higher rates in LMICs[3]. Mehtar, S., et al.[14] and Allegranzi, B., et al[15] agree that SSI's are the most common health-care-associated infections in LMICs with rates ranging from 8 to 30% with substantial morbidity, mortality, and economic impacts.

The Epidemiology Burden

The rate of surgical site infections in low- and middle-income countries (LMICs) varies widely depending on the hospital and type of procedure. In some groups it is only a few percent, while in others it is very high. Studies show that many hospitals have high baseline infection rates, but a large share of these infections could be prevented through quality improvement measures.

Examples of the quality improvement measured rates

A multisite Clean-Cut implementation showed baseline 30-day SSI of 28.4% with reduction to 12.1% after intervention in five low-resource hospitals, absolute reduction 16.3% (RR 0.51)[14]. A prospective cohort in Freetown, Sierra Leone, found an overall SSI incidence of 11.5% (69% of SSI deep) among 338 adults undergoing surgery[15].

A Tanzanian obstetric unit recorded post-caesarean SSI prevalence of 48% pre-intervention and 17% post-intervention in sequential surveys[16]. Therefore, implementing this study could

significantly reduce SSI rates in the Ugandan setting by enabling earlier detection and timely interventions as part of quality improvement projects.

Clinical Consequences

SSIs drive longer hospital stays, readmissions and reoperations and are associated with increased morbidity and mortality; these downstream effects amplify patient and health-system costs in resource-limited settings[13, 17].

2.3 Risk Factors and Early Detection Challenges

Surgical site infections (SSIs) remain among the most common healthcare-associated infections globally and are a major contributor to postoperative morbidity, prolonged hospitalization, and increased healthcare costs, particularly in low- and middle-income countries (LMICs). The identification of well-established risk factors is essential for both clinical risk stratification and the development of predictive models.

2.3.1 Well-Known Clinical and Procedural Risk Factors

Extensive literature has identified several intraoperative and procedural determinants of SSI. One of the most consistently reported risk factors is prolonged duration of surgery. Longer operative times increase tissue exposure, microbial contamination risk, and physiological stress, thereby elevating infection probability[18, 19]. Similarly, the type of surgical procedure particularly emergency and contaminated surgeries has been associated with increased SSI incidence[20].

Antibiotic prophylaxis practices also play a critical role. Inappropriate timing, incorrect antibiotic selection, or delayed administration significantly increase postoperative infection risk and inadequate adherence to infection prevention and control (IPC) protocols during surgery, including inadequate skin preparation, non-adherence to sterile protocols or techniques and wound care management, contributes substantially to infection development in LMICs[21, 22].

Length of hospital stay has also been associated with SSI risk. Prolonged preoperative or postoperative hospitalization increases exposure to nosocomial pathogens and may reflect underlying complications or vulnerability. In maternal populations, Caesarean delivery is a recognised high-risk procedure for SSI, particularly in resource-constrained settings[23, 24].

2.3.2 Patient-Related Risk Factors

Patient-level characteristics are equally important determinants of SSI, advanced maternal age has been associated with impaired wound healing and increased comorbidity burden. Obesity and elevated Body Mass Index (BMI) are strongly linked to higher SSI risk due to poor tissue perfusion and impaired immune response[25].

These factors collectively underscore the multifactorial nature of SSI and justify the inclusion of demographic, clinical, and perioperative features in predictive modelling.

2.3.3 Postoperative Factors

Post-operative wound care practices significantly influence infection outcomes. In LMIC settings, limited access to sterile equipment, inconsistent IPC enforcement, and shortages of trained personnel may increase vulnerability to SSI.

Post-discharge care is particularly critical in maternal populations. Many SSIs manifest after hospital discharge, especially in settings where early discharge is common. Inadequate patient education regarding wound care and limited access to follow-up services further complicate early detection.

2.3.4 Challenges in Early Detection of SSI

Early detection of SSI remains challenging for several reasons. First, early symptoms such as mild redness, swelling, or discomfort often overlap with normal postoperative healing processes, making clinical differentiation difficult. This overlap may delay diagnosis until infection becomes more severe.

Second, delayed manifestation is common. A substantial proportion of SSIs present days or weeks after discharge, particularly following Caesarean delivery. In resource-limited contexts, structured post-discharge surveillance systems are often weak or absent, leading to underreporting and delayed intervention.

Third, detection frequently relies on subjective clinical judgment. Variability in clinician experience, training level, and documentation practices can introduce observer bias and misclassification. Inconsistent diagnostic criteria further contribute to variability in reported SSI rates.

2.3.5 Rationale for Machine Learning in SSI Detection

The multifactorial and nonlinear nature of SSI risk, combined with challenges in early clinical detection, highlights the potential value of machine learning (ML) approaches. Unlike traditional statistical methods that may rely on predefined linear relationships, ML models can identify complex interactions among patient-related, procedural, and operational variables.

ML can support early risk stratification by integrating multiple predictors simultaneously and generating real-time risk scores. In settings where clinical expertise and surveillance infrastructure are limited, predictive models may assist healthcare providers in identifying high-risk patients for closer monitoring.

Given the overlapping symptoms, delayed manifestation, and subjective nature of SSI diagnosis, data-driven predictive systems offer an opportunity to enhance consistency, reduce misclassification, and strengthen infection prevention and control strategies. Therefore, the integration of ML techniques into maternal health surveillance represents a promising avenue for improving early detection and reducing postoperative complications in low-resource settings.

2.4 Machine Learning for Surgical Site Infections

Machine learning (ML) has emerged as a promising tool for predicting surgical site infections (SSIs) following caesarean sections, enhancing patient care and resource allocation. Various studies demonstrate the effectiveness of ML models in identifying risk factors and improving surveillance systems for SSIs, particularly in diverse healthcare settings.

2.4.1 Predictive Models and Performance

In a study conducted in Rwanda, logistic regression and support vector machine (SVM) models achieved high accuracy (AUC = 99.9%) in predicting infections based on questionnaire and image data from caesarean section wounds[26]. Research on colon surgeries showed that XGBoost

outperformed traditional models, achieving an AUC-ROC of 0.788, indicating its potential for effective SSI prediction.

2.4.2 Data Utilization

Structured and Unstructured Data: ML models utilize both structured data (demographics, medications) and unstructured clinical notes, enhancing the predictive capability for SSIs[27]. Advanced models like long short-term memory networks have shown superior sensitivity (92%) compared to traditional methods, indicating their effectiveness in real-time SSI identification [27].

2.4.3 Implications for Healthcare

By accurately predicting SSIs, ML can help prioritise high-risk patients, potentially reducing hospital stays and healthcare costs associated with infections. The integration of ML in SSI surveillance can automate labor-intensive processes, improving efficiency and accuracy in identifying infections[27].

While machine learning shows significant promise in enhancing SSI prediction and management, challenges such as data imbalance and the need for interpretability in models remain critical considerations for future research and implementation.

Al Mamlook et al[28] investigated the application of machine learning to predict surgical site infections (SSIs), a major healthcare challenge due to their contribution to increased morbidity, mortality, and costs.

Using data from the American College of Surgeons' National Surgical Quality Improvement Program (ACS NSQIP), the study evaluated seven different models; Logistic Regression, Naïve Bayes, Random Forest, Decision Tree, Support Vector Machine, Artificial Neural Network, and Deep Neural Network.

Among these, the Deep Neural Network demonstrated the highest predictive performance (AUC = 0.85), significantly outperforming traditional approaches, particularly Logistic Regression, which showed only moderate accuracy.

Key risk factors identified included emergency surgeries, ASA Index of 4, higher BMI, vascular procedures, and general surgeries. Overall, the results from Al Mamlook et al study suggest that utilization modern ML approaches can significantly strengthen SSI prediction models.

ML applications in maternal health are gaining momentum, especially in risk prediction for conditions such as preeclampsia, gestational diabetes, and postpartum hemorrhage. However, the application to maternal sepsis remains underexplored.

A recent study in India applied decision trees and logistic regression to identify early sepsis indicators in postpartum women, demonstrating moderate predictive power with AUC values ranging between 0.72–0.81. Another study in Nigeria used an artificial neural network to detect infection progression, showing improved sensitivity over clinician-based assessments.

Machine learning (ML), a subset of artificial intelligence, offers tools capable of recognizing complex patterns in large clinical datasets. ML models such as logistic regression, K-Nearest Neighbor, support vector machines (SVM), random forests (RF) and gradient boosting.

For instance, studies have demonstrated that ensemble methods like XGBoost and LightGBM outperform traditional statistical models in predicting sepsis using electronic health record (EHR) data[29].

In most cases these are used. While many ML models have shown promise, most research is concentrated in high-income settings, with limited validation in LMICs where contextual differences in clinical practices and data availability exist.

These studies suggest that localized, context-aware ML models may offer valuable tools for improving maternal outcomes in LMICs. However, the concept of explainability of how the ML models came up with the prediction is in black box.

2.5 Explainable Artificial Intelligence (XAI)

One of the main criticisms of ML in healthcare is its "black-box" nature. Clinicians are hesitant to adopt models they cannot understand[30]. Explainable AI (XAI) addresses this by providing insights into how models make predictions. For example:

- Feature importance methods rank variables (e.g., high blood sugar, prolonged labor) by their influence on predictions.
- LIME (Local Interpretable Model-agnostic Explanations) generates human-readable explanations for individual predictions.
- SHAP (SHapley Additive exPlanations) provides consistent attribution of feature contributions across predictions.

In infection-related studies, XAI has been used to highlight clinical risk factors, improving clinician trust and decision-making.

2.6 Machine Learning and Explainability in Low resource setting

Okolo et al. highlight critical gaps in the development and application of explainable AI (XAI) for low-resource settings, particularly in the Global South[31]. Their review reveals that most existing XAI research is rooted in the Global North and assumes a user base that is relatively technology-literate. This overlooks the realities of low-income and low-literate populations, where limited digital literacy significantly influences technology adoption.

They argue that explanations must be tailored to the intended audience, which may include both system developers and end users such as frontline health workers, many of whom have little prior experience with AI systems. The authors call for human-centered XAI approaches that adapt explanations into accessible languages, culturally appropriate formats, and computationally feasible designs for low-resource environments.

Importantly, Okolo et al. caution against "explainability pitfalls," where users may place blind trust in AI outputs simply because an explanation is provided. This risk is particularly pronounced in low-resource contexts, where users might defer to AI-generated explanations even when the underlying prediction is wrong due to limited technical expertise. For example, a community health worker may accept an AI's incorrect pneumonia diagnosis if the explanation appears authoritative.

Their work stresses the need for future XAI research that not only improves technical interpretability but also addresses social, cultural, and ethical challenges in low-resource settings. By doing so, XAI can empower rather than mislead users, ensuring that explainability becomes a tool for equitable and trustworthy AI deployment in the Global South.

2.7 Explainable Artificial Intelligence (XAI) in Medical Care

Antoniadi et al.’s highlights the urgent need for XAI in medical care to address bias, support fair and ethical decision-making, and align AI-driven systems with human rights and regulatory frameworks. XAI offers a vital mechanism for exposing bias, fostering accountability, and guiding the development of responsible AI in healthcare.

The integration of artificial intelligence (AI) into clinical decision-making has highlighted a pressing need for transparency and accountability. Antoniadi et al. emphasize that “black box” AI models, which provide predictions without explanations, pose significant risks due to their opacity and tendency to conceal systemic biases[32]. When left unexamined, these biases can reinforce pre-existing societal inequities, disproportionately disadvantaging already marginalized groups.

Explainable AI (XAI) emerges as a potential solution to these challenges by enabling clinicians and patients to interrogate model outputs, trace decision pathways, and assess fairness. However, Anna et al. caution that existing XAI techniques remain limited in their ability to expose discriminatory behaviours in sensitive applications. Moreover, oversimplified explanations risk misleading users, providing a false sense of trust, or even facilitating “fair-washing,” whereby models appear ethically sound while masking underlying inequities. XAI the ability should enhance a model’s ability to generate explanations that genuinely support human understanding and ethical use in practice.

Building on this discussion, Pawar et al. (2020)[33] stress the practical value of XAI in strengthening medical practitioners’ confidence in AI-based solutions and improving their understanding of model operations. Their work introduces a feature importance (FI)-driven approach to explanation, highlighting how FI scores can reveal the relationships between features, models, and data within clinical workflows.

They outline three complementary perspectives: (i) a model/output-based view, which identifies features most influential to model outcomes; (ii) a hierarchical ranking across medical settings, which illustrates how feature importance may vary depending on diagnostic contexts, thereby aligning AI explanations with real-world practice; and (iii) the impact of data augmentation, which demonstrates how synthetic data affects model performance and the validity of FI scores in medical contexts.

Collectively, these approaches allow practitioners to assess not only which features drive predictions but also how models may behave under different conditions, ultimately enabling more informed and trustworthy decision-making. Importantly, Pawar et al. propose extending this work through a multi-level framework that automates model training and validation, with the aim of embedding explainability across hierarchical layers of medical decision-making.

This perspective underscores XAI’s dual role: not only as a tool for transparency and fairness, as argued by Antoniadi et al., but also as a means of operational efficacy, helping practitioners better interpret, validate, and apply AI solutions in diverse clinical contexts.

Together, the contributions of Antoniadi et al. and Pawar et al. illustrate that the need for XAI in healthcare extends beyond addressing ethical and legal obligations; it is also central to fostering clinical trust, interpretability, and practical utility of AI-driven systems.

2.8 Explainability and ML Based Predictive Modelling

Explainability and predictive performance are related but not strictly inverse: empirical studies show the relationship depends on model class, data, and the human evaluation criteria used, so a single “accuracy vs interpretability” rule does not hold universally[34, 35].

When a trade-off appears, it stems from (a) limiting model capacity to force transparency, (b) using post-hoc explanations that may lack faithfulness, and (c) human cognitive limits that make more information confusing rather than clarifying[36, 37, 38].

Traditional black-box models often capture richer patterns that can increase accuracy; constraining models for transparency can reduce capacity and thus performance in some settings[38].

Post-hoc explanations (surrogates, attribution methods) can be informative but may not faithfully reflect the true model decision boundary, creating a separate fidelity-versus-interpretability trade-off[39].

More explanation content can overload users and reduce comprehensibility; perceived explainability depends on user task and presentation, not only model structure[35, 40].

Practical strategies balance objectives by matching explanation style to purpose, validating explanation fidelity, and iterating with users; these approaches reduce unnecessary performance loss while improving trust[40]. Future work should standardize fidelity metrics, expand human-centered evaluations, and develop methods that produce guaranteed faithful, compact explanations without degrading predictive power.

2.8.1 Research priorities and open problems

Algorithms that guarantee meaningful explanation without sacrificing performance are lacking in literature[41, 42].

2.9 Research Gaps Identified

Despite increasing interest in the application of machine learning in healthcare, significant methodological and contextual gaps remain in the prediction of postpartum surgical site infections (SSIs), particularly in low- and middle-income countries (LMICs). Existing studies on SSI prediction predominantly rely on traditional statistical approaches, such as logistic regression, with limited adoption of advanced machine learning techniques for early risk stratification[2, 13].

The integration of explainability frameworks such as SHAP or LIME remains limited in SSI prediction studies, despite the importance of model transparency for clinical decision-making in LMIC maternal health settings. Where explainability approaches are employed, they are predominantly applied in high-income country contexts, thereby limiting contextual relevance and transferability to Sub-Saharan African health systems[43, 44].

There is also a notable scarcity of locally validated models developed using large, context-specific datasets from Sub-Saharan Africa. Many predictive models are developed and validated in high-income countries, where differences in epidemiological patterns, healthcare infrastructure, perioperative practices, and post-discharge follow-up systems may limit their generalizability to LMIC contexts[1, 45].

Furthermore, studies conducted in LMIC settings often rely on relatively small datasets, which constrains model robustness, external validity, and reproducibility. Ultimately, these gaps

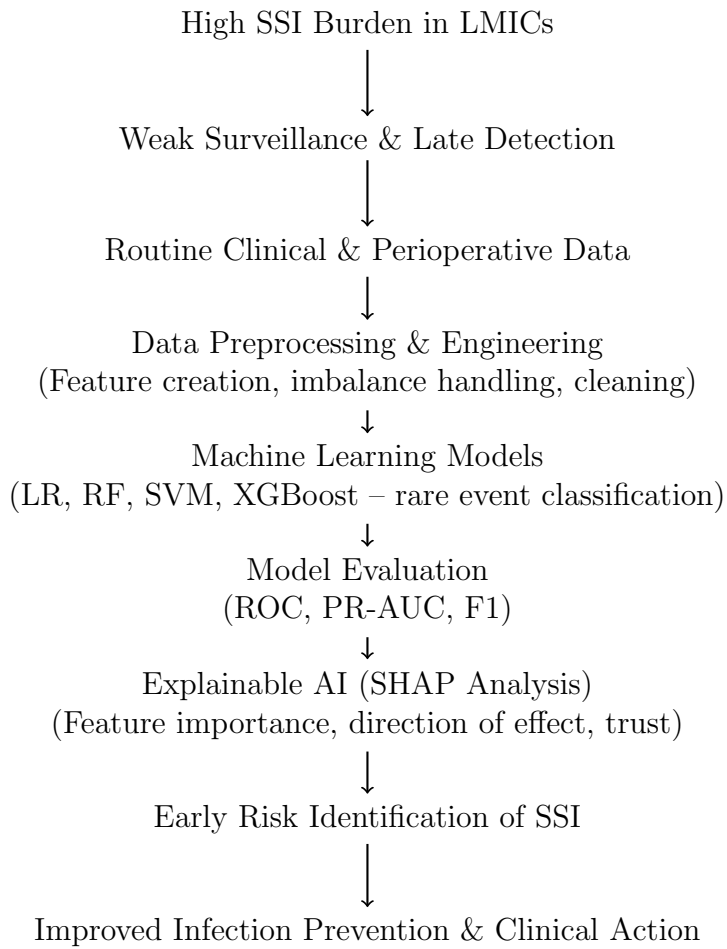
underscore the need for robust, explainable, and context-sensitive machine learning frameworks for early detection of postpartum SSIs in Uganda.

Therefore, this study aims to address these gaps by developing and applying explainable machine learning models for the early detection of postpartum SSIs in Ugandan healthcare settings, using a dataset drawn from multiple sites across three different regions.

Table 2.1: Research Gap Summary

Author (Year)	Country/Context	Study Focus	Methodology	Key Findings	Relevance to Current Study	Gap Identified
Badia et al. (2017)	Europe	Economic impact of SSI	Systematic review	SSIs increase length of stay and costs	Establishes importance of early detection	Limited LMIC evidence
Chu et al. (2015)	Sub-Saharan Africa	Post-C-section SSI rates	Multi-country observational	High SSI prevalence in SSA	Confirms burden in similar context	No predictive modelling
Wood et al. (2025)	SSA	SSI prevalence and factors	Scoping review	Wide prevalence variation; surveillance gaps	Justifies surveillance strengthening	No ML application
Mehtar et al. (2020)	LMICs	SSI surveillance systems	Position statement	SSI rates 8-30% in LMICs	Supports need for structured monitoring	No predictive analytics
Nofal et al. (2025)	5 LMIC hospitals	Clean-Cut intervention	Multisite implementation	SSI reduced from 28.4% to 12.1%	Shows improvement via structured prevention	No early risk prediction
Lakoh et al. (2022)	Sierra Leone	SSI incidence	Prospective cohort	11.5% SSI incidence	Demonstrates comparable burden	No ML
Gentilotti et al. (2020)	Tanzania	IPC + stewardship	Intervention study	SSI reduced from 48% to 17%	Highlights modifiable risk factors	No predictive modelling
Fletcher et al. (2019)	Rwanda	ML for SSI	Logistic Regression, SVM	AUC = 0.999 using image + survey data	Demonstrates ML feasibility in Africa	Limited generalizability
Al Mamlook et al. (2023)	USA	ML SSI prediction	7-model comparison	DNN best (AUC = 0.85)	Shows ML improves SSI prediction	High-income dataset
Kiser et al. (2024)	USA	LSTM for SSI	Deep learning	Sensitivity 92%	Demonstrates sequential modelling	Resource-intensive
Islam et al. (2023)	Global	ML for sepsis	Systematic review	Ensemble models outperform traditional	Supports infection prediction via ML	Limited postpartum focus
Okolo et al. (2022)	Global South	XAI in low-resource settings	Systematic review	Need for human-centered XAI	Supports contextual explainability	Limited clinical implementation
Antoniadi et al. (2021)	Global	XAI in healthcare	Systematic review	Black-box models risk bias	Justifies SHAP integration	Limited LMIC validation
Pawar et al. (2020)	Healthcare	Feature importance methods	Framework proposal	FI improves trust	Supports SHAP use	Limited empirical LMIC testing
Kruschel et al. (2025)	Multi-domain	Accuracy vs explainability	Empirical study	Trade-off context dependent	Supports balanced model selection	Not SSI-specific

2.10 Conceptual Framework



Chapter 3

Methodology

3.1 Introduction

This study aimed to assess the feasibility of using explainable machine learning models for the early detection of postpartum Surgical Site Infections (SSI) based on clinical and demographic data. This was achieved through the development and evaluation of multiple supervised machine learning algorithms, with a focus on model interpretability and transparency.

In addition, the study sought to identify gaps in routine hospital data that may limit predictive performance with the goal to inform both clinical decision-making and improvements in data collection practices to enhance early detection and intervention.

3.2 Study Design

This study adopted a quantitative and retrospective predictive design, guided by the data science lifecycle. The research method followed five phases: data acquisition, preprocessing, analysis, modelling, and evaluation.

Data pre-processing included handling missing values, outliers, and inconsistencies, followed by exploratory data analysis (EDA) to uncover trends, distributions, and relationships among patient characteristics. Feature engineering was applied to derive additional variables that may enhance predictive performance.

Supervised machine learning algorithms, including logistic regression, decision trees, random forests, support vector machines, and gradient boosting models, were implemented to predict the likelihood of postpartum surgical site infections. Performance was evaluated using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC. To enhance interpretability and clinical relevance, SHAP Explainable AI (XAI) technique was applied to clarify how input features influence model predictions.

3.3 Data Source

3.3.1 Data Collection and Acquisition

The dataset used in this study was obtained from the Surgical Site Infection (SSI) surveillance under the Healthcare-Associated Infections (HAI) study conducted by the Infectious Diseases

Institute. This dataset constitutes secondary data, as it was originally collected for routine clinical surveillance rather than specifically for this research.

Data were collected prospectively between May 2024 and January 2026 from three Regional Referral Hospitals: Kayunga, Hoima, and Entebbe. The data collection process utilized the Open Data Kit (ODK) platform, with information captured electronically using one primary data collection tool at the time of surgery and three follow-up tools administered on post-operative days 6, 14, and 30 to monitor patient outcomes, particularly the occurrence of surgical site infections.

All data were stored centrally on the ODK server managed by the Infectious Diseases Institute. For the purpose of this study, the researcher developed a Python-based data processing pipeline to extract the four datasets (baseline and follow-up forms), clean them, and merge them into a single analytical dataset. The merging process was performed using unique patient identification numbers as the primary key to ensure accurate linkage of patient records across all follow-up time points.

3.3.2 Dataset Description

This data set contained more than 8,900 surgical cases and 34 features recorded during the Healthcare-Associated Infection (HAI) Surveillance Program between 20th May 2024 and 17th January 2026. Each record represents a unique patient ID, both for the study and a local identifier, and includes demographic and clinical variables. The target variable, `final_outcome`, is binary and indicates whether the patient had an infection or not (1 = infection present, 0 = no infection) at day 30 after discharge.

The data set contained 34 features: Date of Admission, Patient_ID, study site, Referral In, age, c_section_type, indication_of_c_section, c_section_date, start_time_of_surgery, end_time_of_surgery, surgery_by, anesthesia_by, asa_score, HIV, did_patient_shower, hair_removal_done, patient_skin_preparation, Did_the_skin_dry_before_the_incision_was_made, Was_blood_glucose_done, Was_antibiotic_Prophylaxis_given, Was_the_patient_taught_about_hygiene_and_wound_care_after_surgery, Were_antibiotics_given_post_operati, purulent_drainage, wound_reopened_first_check, fever, fever_and_wound_redness_first_check, fever_and_woun, fever_and_increased_wound_pain, date_of_review_discharge, date_antibiotic_given, number_of_times_adminis, number_of_days_given, Purulent_Drainage_Abscess_Day_30_post_discharge, final_outcome.

Item	Description
Source	HAI Surveillance Database
Collection Tool	ODK
Sites	Hoima, Kayunga, Entebbe RRH
Study Period	May 2024–Jan 2026
Records Extracted	8,918
Outcome	SSI at Day 30

Table 3.1: Summary of Dataset Characteristics

3.3.3 Data Governance and Inclusion Criteria

Data originated from the Healthcare Associated Infection (HAI) surveillance programme implemented by the Infectious Diseases Institute in collaboration with participating Regional Referral Hospitals. Eligible records included women undergoing Caesarean section with complete baseline surgical information and at least one documented follow-up outcome (day 6, 14,

or 30). Duplicate records were removed using patient identifiers and outcome reconciliation rules. Variables with excessive missingness > 70% were excluded from modelling

3.4 Data Management Plan

Data management procedures were implemented to ensure integrity, confidentiality, reproducibility, and secure handling throughout the research lifecycle. Raw data extracted from ODK servers were stored on password-protected encrypted devices accessible only to authorized researchers. Data cleaning and transformation steps were implemented using version-controlled Python scripts to preserve reproducibility. Direct identifiers were removed before analysis and only de-identified analytical datasets were retained. Final datasets, scripts, and outputs were archived following institutional research data governance procedures.

3.5 Data Cleaning and Preprocessing

Data pre-processing was undertaken to enhance the quality, consistency, and suitability of the dataset for machine learning modelling. The initial step involved data cleaning, where missing values were addressed using appropriate imputation techniques, including Multiple Imputation by Chained Equations (MICE). Data anomalies such as negative or implausible values were identified and corrected, with invalid entries replaced with logical defaults.

Categorical variables were transformed into numerical representations to facilitate model compatibility. Depending on the nature of the variable and the requirements of specific algorithms, encoding techniques such as ordinal encoding and one-hot encoding were applied.

To ensure comparability across features, numerical variables were scaled using standardization techniques, particularly the StandardScaler, which transforms features to have a mean of zero and a standard deviation of one. This step was especially important for distance-based and gradient-based models.

The purpose of data cleaning and pre-processing was to develop a thorough understanding of the dataset, particularly the different features, data types, and the presence of missing or inconsistent data. This initial examination also allowed for the identification of outliers and potential errors that could affect subsequent analyses. Insights from the exploration informed the need for data pre-processing, including the conversion of data types, handling missing values, and the creation of new derived variables to facilitate analysis.

The dataset comprised participants who underwent Caesarean section (C-section) procedures over a one-year period, from 20th May 2024 to 17th January 2026, across three facilities: Kayunga, Hoima, and Entebbe Regional Referral Hospitals. For consistency, the total number of participants with surgical site infections (SSI) was computed from follow-up-specific data collected at baseline, Day 6, Day 14, and Day 30. In cases where participants had multiple SSI entries due to data entry errors, the SSI outcome recorded at the first follow-up date was considered.

Participants who died during the surveillance period were excluded from the analysis to ensure data consistency. Additionally, date variables such as Admission, C-section, first check, and Discharge were converted to proper date-time formats to enable accurate computation and analysis.

3.5.1 Missing Data Analysis

Missing data analysis was conducted as part of the data cleaning and preprocessing stage to identify the presence, patterns, and mechanisms of missing values within the dataset. Visual exploration techniques were employed using the Missingno library in Python.

A missingness matrix was first generated to visualize the distribution of missing values across the dataset, allowing the identification of variables with incomplete observations. This visualization helped highlight variables with missing entries and provided an overview of the completeness of the dataset.

To further investigate relationships between missing values across variables, a missingness correlation heat map was produced. The heat map revealed that several postoperative clinical variables, including wound condition indicators and follow-up assessments, exhibited strong correlations in their missingness patterns.

This suggests that missing values in these variables tend to occur simultaneously, likely due to dependencies such as missed follow-up visits or incomplete clinical assessments. Additionally, Little's MCAR test was performed to determine whether the missing data were Missing Completely At Random (MCAR).

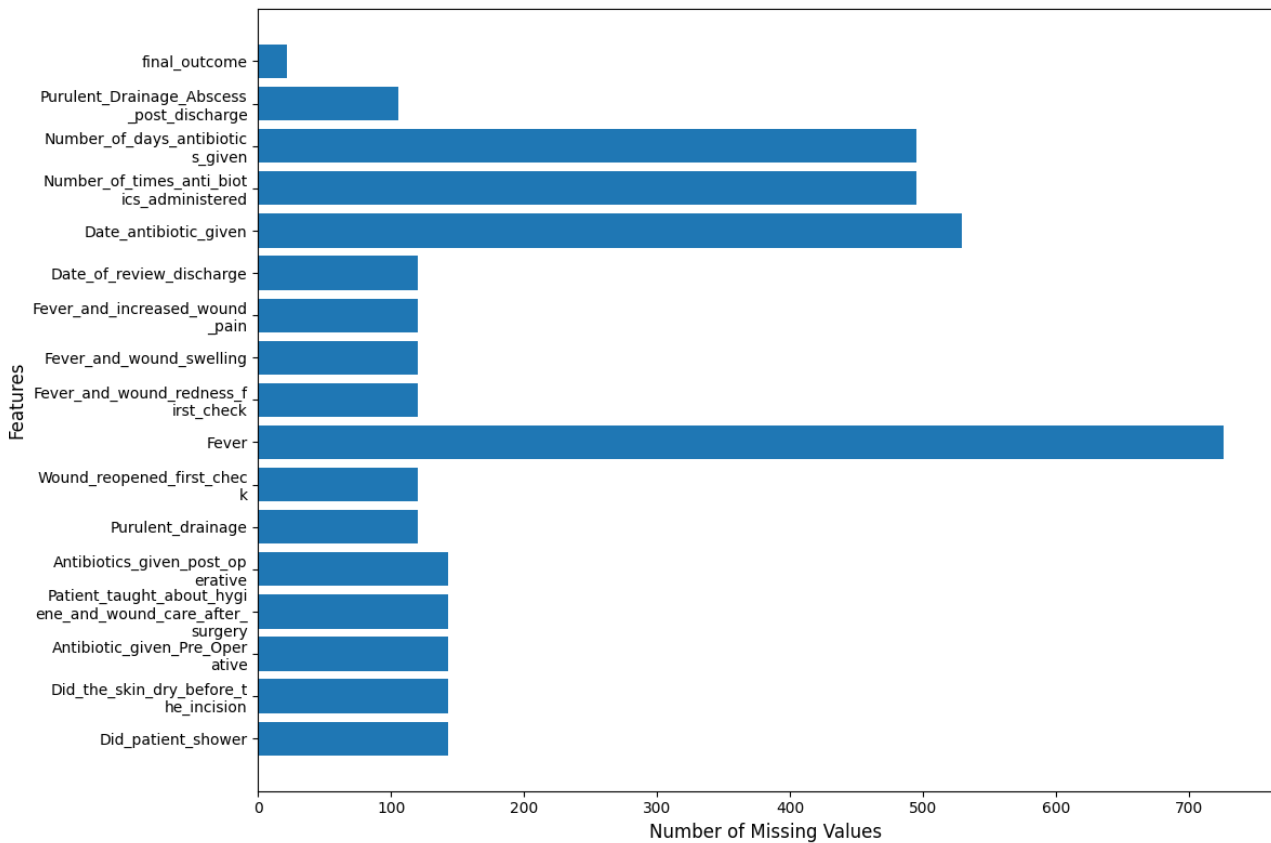


Figure 3.1: Number of missing values per feature in the dataset.

Figure 3.1 presents a horizontal bar chart showing the number of missing values across the dataset's features. The x-axis represents the count of missing observations, while the y-axis lists the corresponding variables.

The variable **fever** exhibits the highest level of missingness, with more than 700 missing observations. Other variables with relatively high levels of missing data include **date_antibiotic_given**,

`number_of_days_given`, and `number_of_times_antibiotic_administered`, each containing approximately 500 or more missing entries. The high proportion of missing values in these variables may be attributed to incomplete clinical documentation or variations in postoperative follow-up procedures.

In contrast, several clinical assessment indicators such as `fever_and_wound_redness_first_check`, `fever_and_increased_wound_pain`, `fever_and_wound_swelling`, `purulent_drainage`, and other related post-discharge monitoring variables demonstrate comparatively lower levels of missingness, generally below 150 observations. This suggests that these symptoms were more consistently recorded during patient follow-up assessments.

The presence of missing data in clinically relevant variables, particularly those related to infection indicators and treatment administration, has the potential to introduce bias and reduce statistical power if not properly addressed.

Consequently, appropriate preprocessing strategies were applied to handle missing values prior to model development. Given that the proportion of missing data was relatively small in most variables, records containing missing values were removed to maintain data quality and ensure reliable model training.

Table 3.2: Summary of Missing Values per Feature

Variable	Missing Values	% of Total Values
Fever	726	8.1
Date_antibiotic_given	529	5.9
Number_of_times_anti_biotics_administered	495	5.6
Number_of_days_antibiotics_given	495	5.6
Did_patient_shower	143	1.6
Antibiotics_given_post_operative	143	1.6
Did_the_skin_dry_before_the_incision	143	1.6
Patient_taught_about_hygiene_and_wound_care_after_surgery	143	1.6
Antibiotic_Prophylaxis_given_Pre_Operative	143	1.6
Fever_and_wound_redness_first_check	120	1.3
Wound_reopened_first_check	120	1.3
Purulent_drainage	120	1.3
Fever_and_wound_swelling	120	1.3
Date_of_review_discharge	120	1.3
Fever_and_increased_wound_pain	120	1.3
Purulent_Drainage_Abscess_post_discharge	105	1.2
final_outcome	22	0.2

The table above presents the distribution of missing values across the variables included in the surgical site infection (SSI) dataset. The analysis indicates that several variables contain missing observations, although the overall level of missingness is relatively low across the dataset.

The variable **Fever** exhibits the highest proportion of missing data, with 726 observations (8.1%) missing. Other variables with notable levels of missingness include `Date_antibiotic_given` (5.9%), `Number_of_times_anti_biotics_administered` (5.6%), and `Number_of_days_antibiotics_given` (5.6%). These variables are related to antibiotic administration and postoperative monitoring, suggesting that some treatment-related information may not have been consistently documented during patient follow-up.

Several clinical assessment variables, including `Did_patient_shower`, `Antibiotics_given_post_operative`, `Did_the_skin_dry_before_the_incision`, `Patient_taught_about_hygiene_and_wound_care_after_surgery`, and `Antibiotic_Prophylaxis_given_Pre_Operative`, each exhibit 1.6% missing values. Similarly, postoperative infection indicators such as `Fever_and_wound_redness_first_check`, `Wound_reopened`

Purulent_drainage, Fever_and_wound_swelling, Date_of_review_discharge, and Fever_and_increased show 1.3% missingness.

The outcome variable `final_outcome` contains 22 missing values (0.2%), representing the lowest level of missingness in the dataset. The relatively small proportion of missing outcome data suggests minimal risk of bias related to the target variable used for predictive modelling.

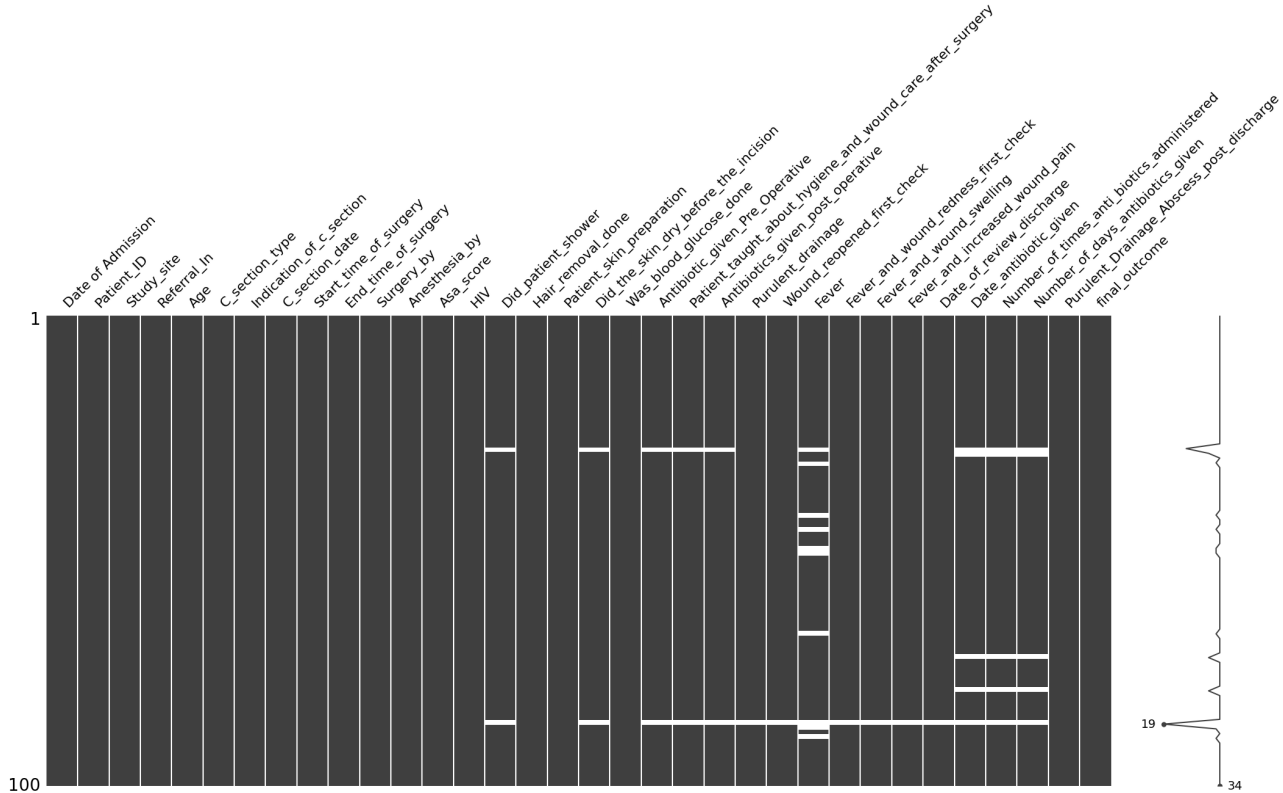


Figure 3.2: Missing data matrix illustrating the pattern of missingness across all variables in the dataset.

Assessment of Missing Data Mechanism

Little's MCAR test was conducted to evaluate the mechanism of missing data. The test produced a p-value of 0.0257, which is less than the significance level of 0.05, indicating that the missing data are not Missing Completely At Random (MCAR). Therefore, the missingness may depend on observed variables, suggesting a Missing At Random (MAR) mechanism where the probability of missingness is related to observed variables rather than occurring entirely randomly.

As a result, appropriate imputation techniques were considered for handling missing values rather than simple row deletion, in order to preserve the integrity and statistical power of the dataset.

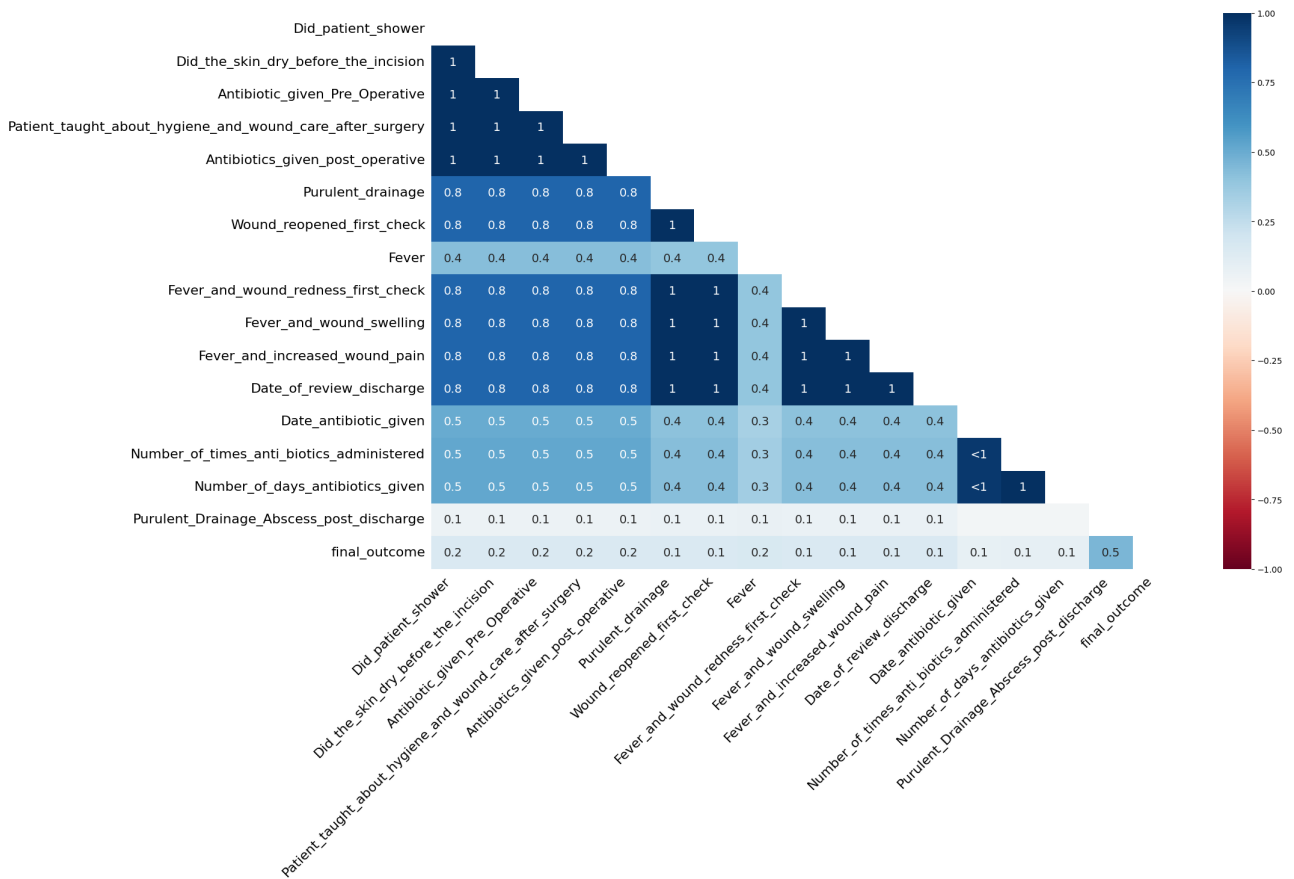


Figure 3.3: Correlation heatmap showing the pairwise relationships among variables with missing data in the surgical site infection (SSI) dataset.



Figure 3.4: Hierarchical clustering dendrogram illustrating similarity relationships among variables in the surgical site infection (SSI) dataset.

Assessment of Missing Data Patterns

Missing data assessment was performed using missingness matrices, summary statistics, and Little's MCAR test. Variables identified as MAR were imputed using MICE.

A correlation heat map was drawn for missing data assessment, darker colours represent stronger correlations, while lighter colours indicate weaker or negligible relationships.

A missing data dendrogram was generated to examine hierarchical clustering of missingness patterns across variables.

3.5.2 Duplicate data

Duplicate data results in inaccurate information and a lack of data integrity. It may result in inaccurate computations and conclusions based on inaccurate information. This particular data set had no duplicates.

3.6 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to gain a comprehensive understanding of the dataset's structure, underlying patterns, and relationships among variables prior to model development. This process was implemented using Python-based data analysis and visualization libraries, including Pandas for data manipulation, and Matplotlib and Seaborn for graphical exploration.

Descriptive statistics were first generated to summarize key characteristics of the data, such as central tendency, dispersion, and distribution. Visual exploration techniques, including histograms and boxplots, were employed to assess the distribution of numerical variables and to identify the presence of outliers and skewness. Scatter plots and pairwise visualizations were further used to examine relationships between predictor variables and the outcome variable, particularly in relation to Surgical Site Infections (SSI).

To evaluate the strength and direction of relationships among variables, correlation analysis was performed. A correlation matrix, complemented by heat maps, was used to visualize linear associations between features and to detect potential multicollinearity. Identifying highly correlated variables informed feature selection and helped reduce redundancy, thereby improving model stability and interpretability.

3.6.1 Univariate Analysis

Univariate analysis was conducted to examine the distribution of individual variables and identify potential data quality issues such as outliers and skewness. Histograms and box plots were generated for numerical variables including age, ASA score, number of times antibiotics were administered, and duration of antibiotic treatment. Each subplot displays the frequency distribution and kernel density estimate, highlighting skewness, outliers, and overall distribution shape of the variables.

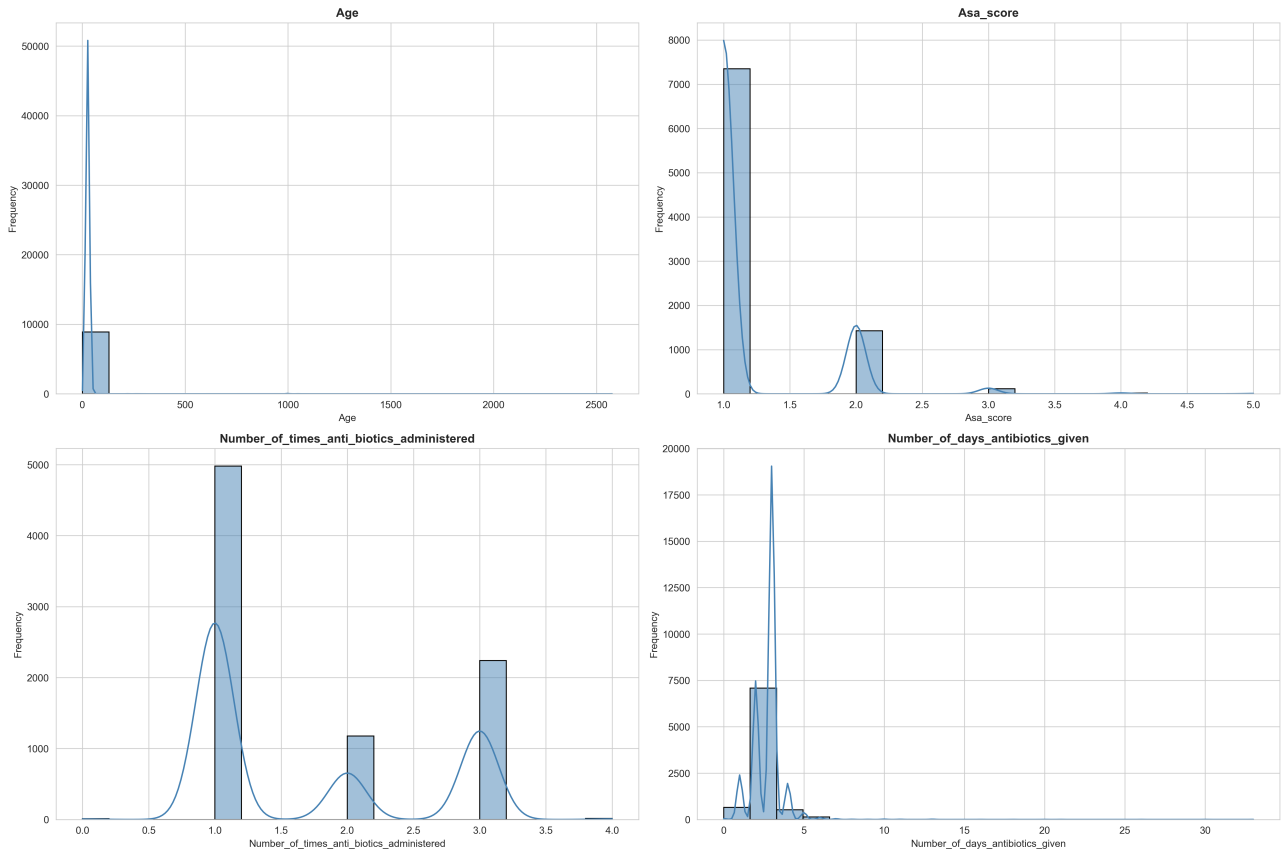


Figure 3.5: Distribution plots for key numerical variables in the dataset.

Univariate Analysis of Numerical Variables

Univariate data analysis revealed several distributional characteristics and potential data quality concerns. The **Age** variable exhibited extreme values exceeding 2000 years, indicating potential data entry or unit inconsistencies, which required verification and cleaning.

The **Asa_score** variable demonstrated a categorical distribution concentrated around lower scores, reflecting that most patients were relatively healthy prior to surgery. Antibiotic administration variables showed clustered distributions corresponding to typical clinical treatment protocols.

The analysis further revealed skewed distributions in several variables and highlighted potential outliers, particularly in the **Age** variable. These visualizations highlight the distribution, central tendency, and presence of outliers across the variables.

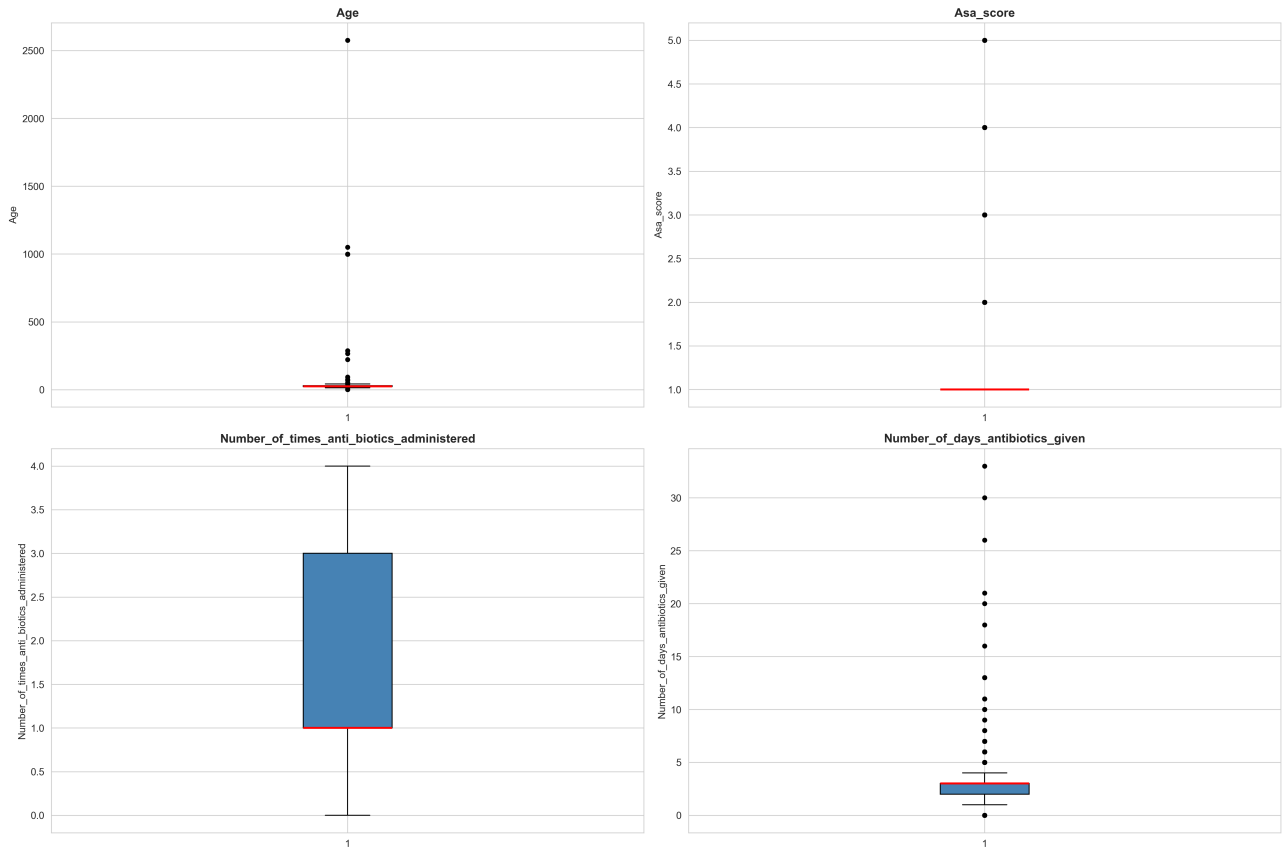


Figure 3.6: Boxplots of selected numerical variables.

Handling the outliers

Outliers were addressed using a two-step approach consistent with best practices in clinical data analysis. First, implausible values arising from likely data entry errors were removed based on clinically acceptable ranges. Second, extreme but plausible values were handled using winsorization, where observations above the 95th percentile were capped at the percentile threshold. This approach preserves the integrity of the dataset while reducing the influence of extreme values on model training.

Dealing with implausible values and winsorization of extreme but plausible values, improved distributional symmetry and reduced excessive dispersion across all continuous variables. While some degree of right skewness persisted in variables, this is consistent with clinical data patterns the cleaned dataset demonstrates more stable central tendencies and reduced influence of extreme values. This ensures that the dataset remains clinically relevant and improves the validity of subsequent analyses. The refined boxplots show reduced extreme values, improved scale visibility, and clearer representation of central tendency and variability following outlier correction.

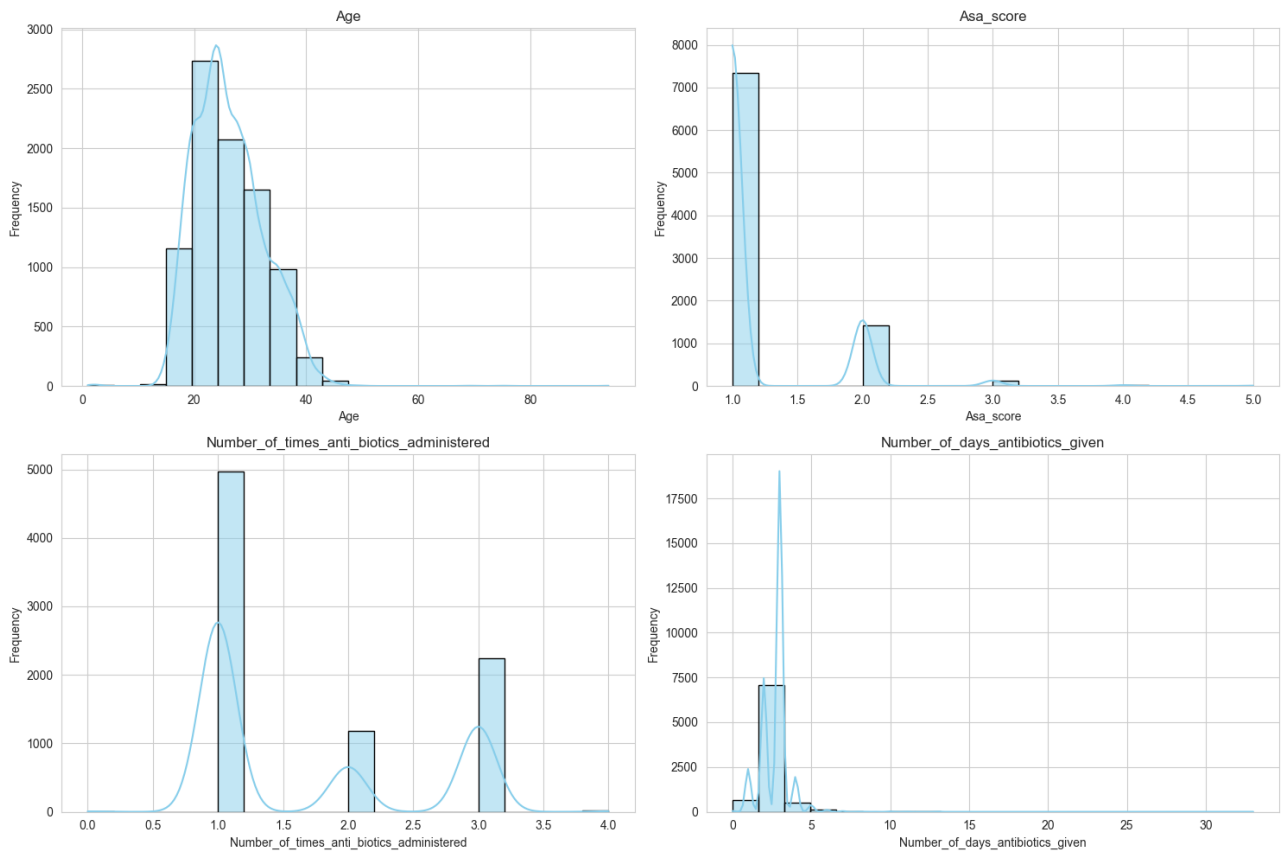


Figure 3.7: Updated distribution plots for key numerical variables after handling outliers.

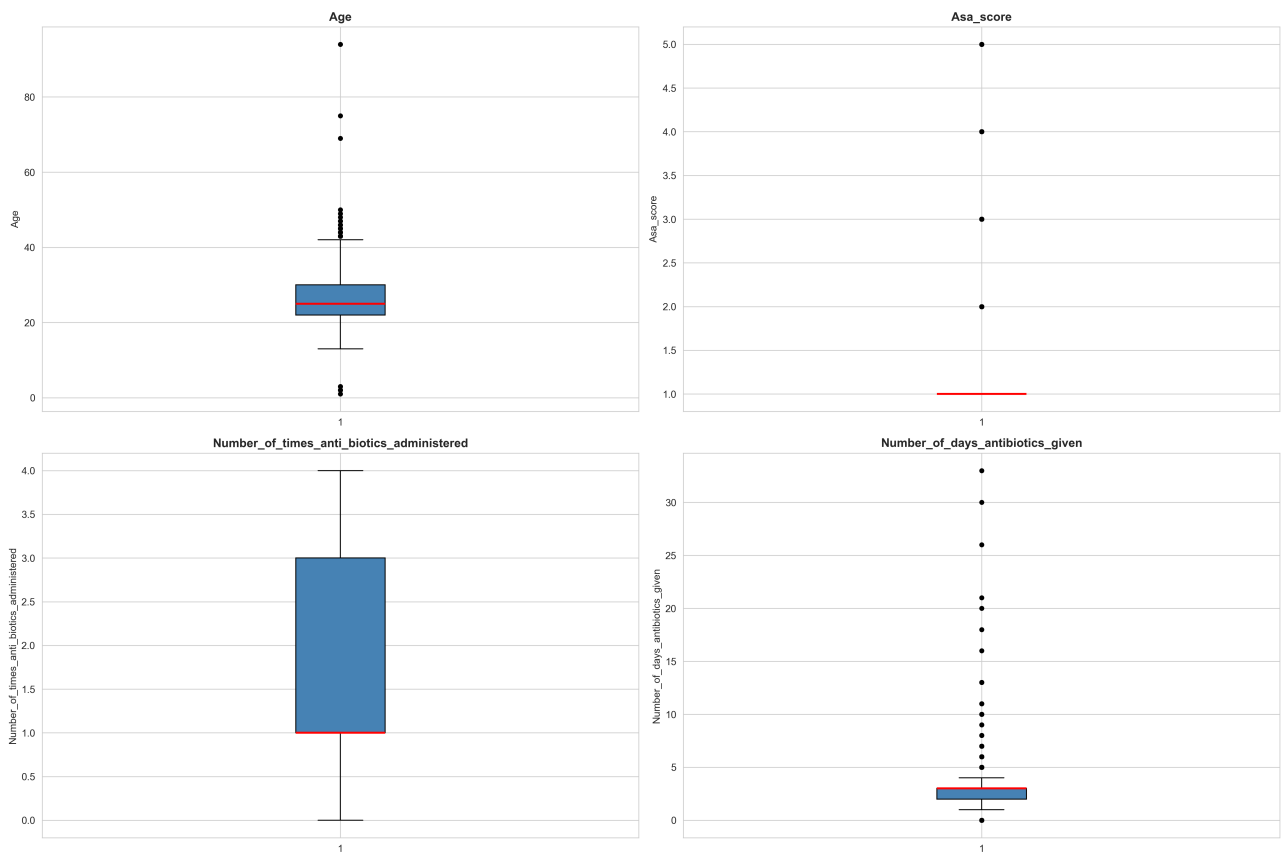


Figure 3.8: Updated boxplots for key numerical variables after handling outliers.

3.6.2 Correlation Matrix

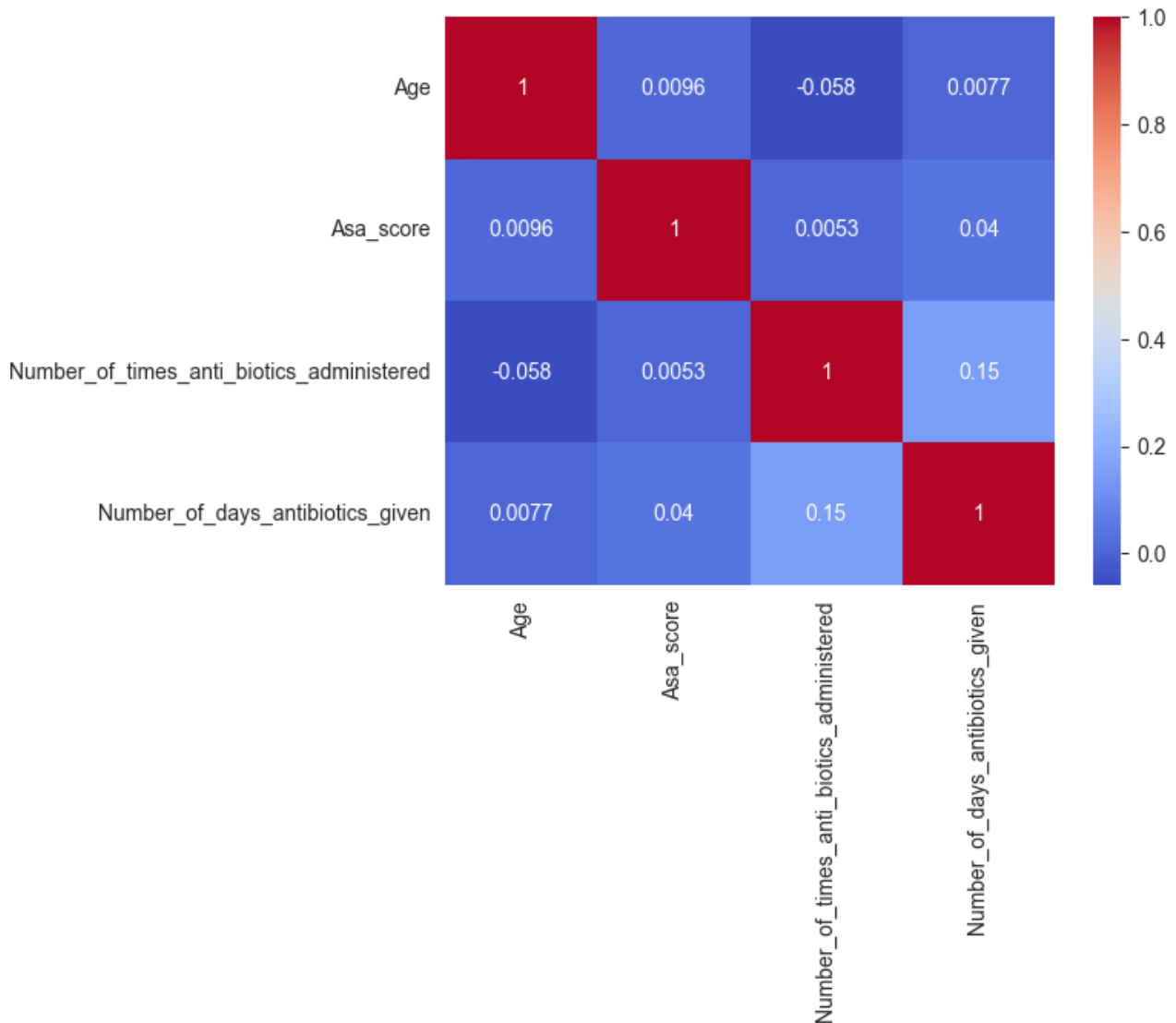


Figure 3.9: Correlation heatmap illustrating pairwise relationships among the cleaned numerical variables

A Pearson correlation analysis was conducted to assess the linear relationships among the continuous study variables, including Age, Asa_score, Number_of_times_anti_biotics_administered, and Number_of_days_antibiotics_given.

Overall, the correlation coefficients were very low, with most values close to zero ($|r| < 0.15$), indicating weak or negligible linear relationships among the variables. The strongest relationship observed was a small positive correlation between the number of times antibiotics were administered per day and the number of days antibiotics were given ($r = 0.15$).

3.6.3 Univariate analysis of Categorical Variables

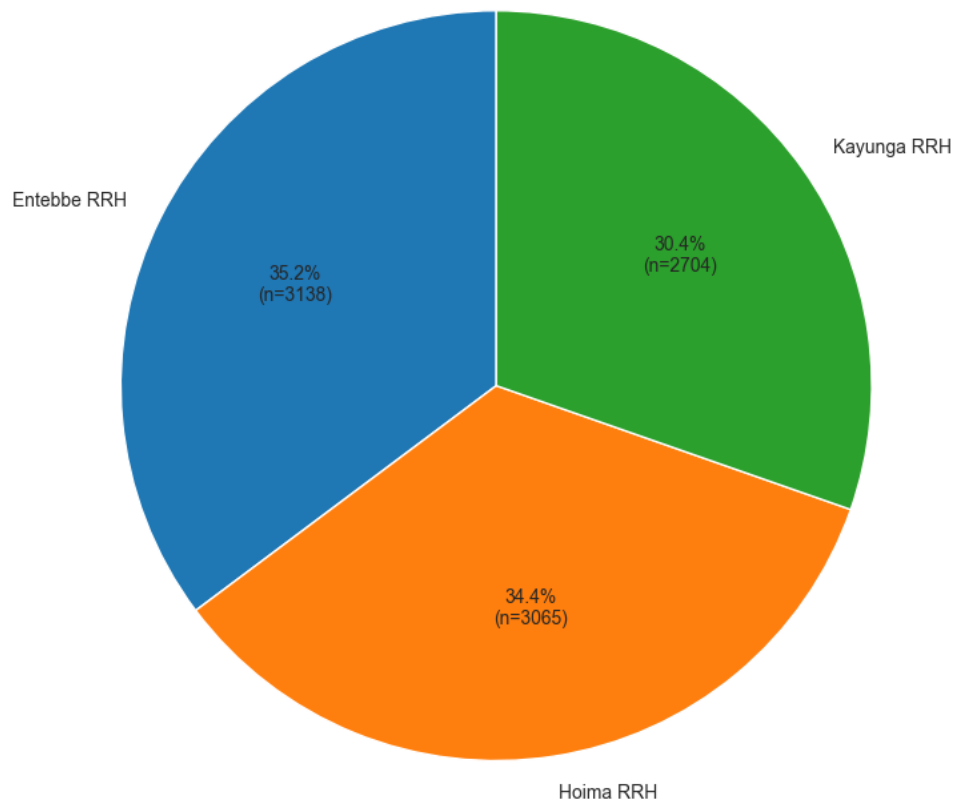


Figure 3.10: Distribution of surgical cases across study sites.

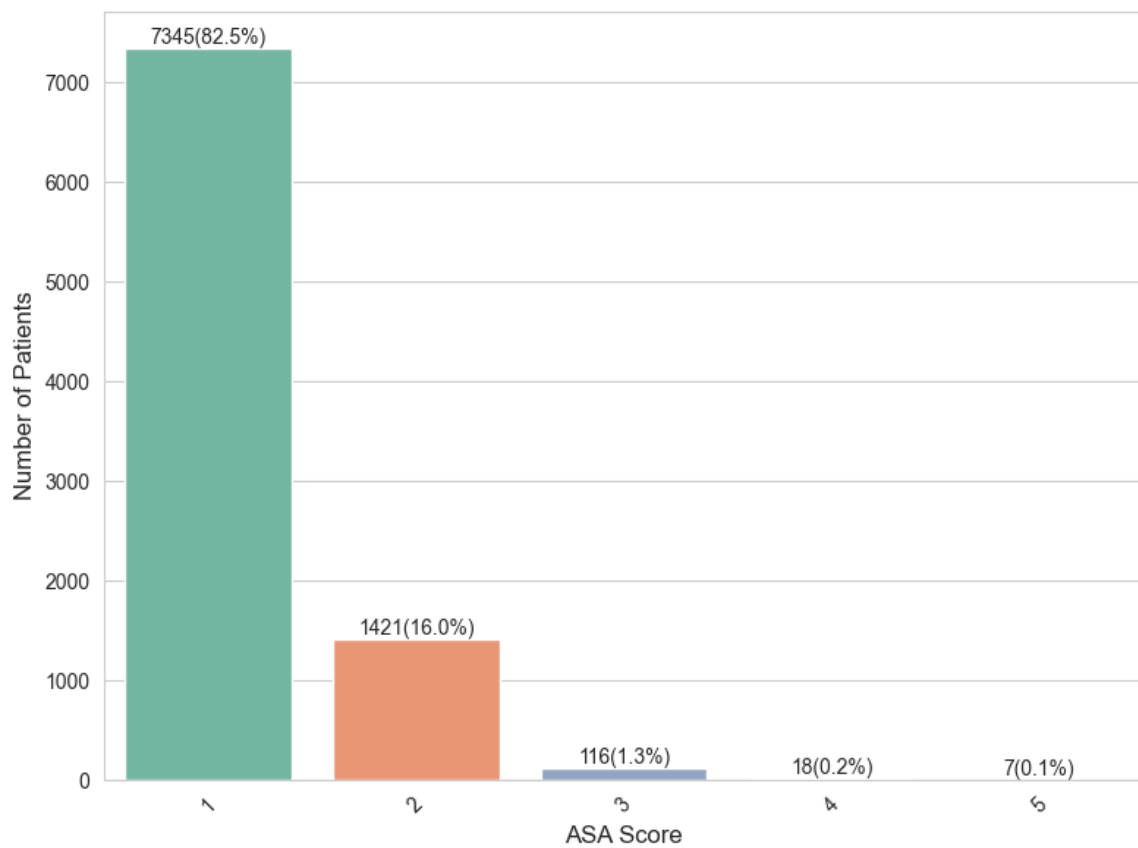


Figure 3.11: Distribution of ASA scores among patients.

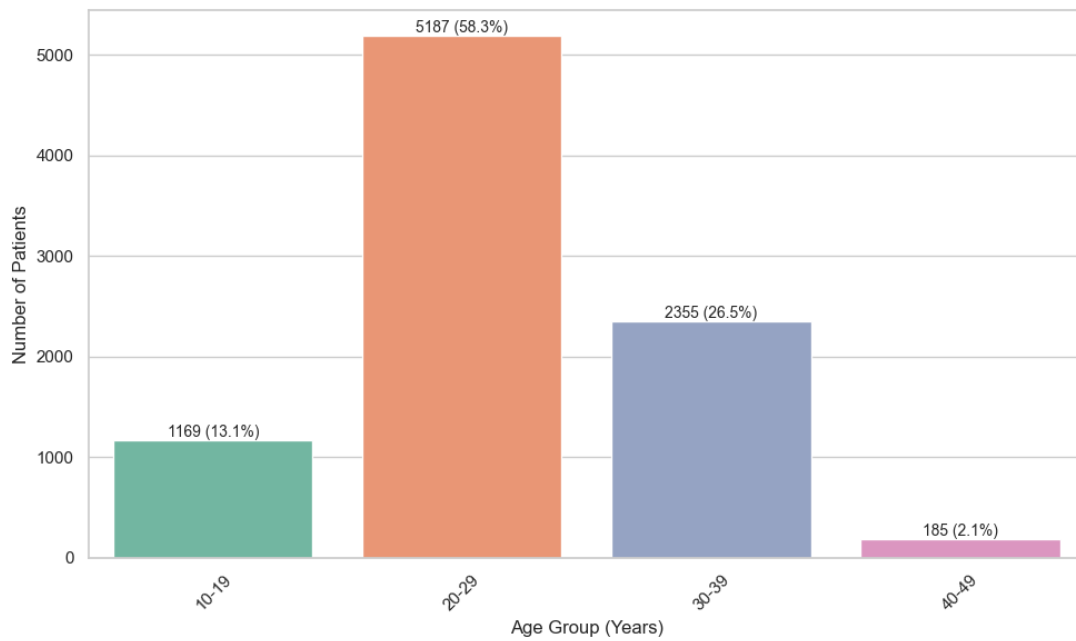


Figure 3.12: Distribution of patients across Age-groups.

3.6.3.1 Distribution of C-Section Type

The analysis revealed that the majority of Caesarean sections performed in the study population were emergency procedures, accounting for 88.9% of cases, while elective Caesarean sections constituted 11.1% of the total procedures.

This indicates that nearly nine out of every ten Caesarean deliveries were performed under emergent circumstances rather than being scheduled in advance. The predominance of emergency procedures may reflect late presentation, obstetric complications, referral patterns, or limited opportunities for planned surgical delivery within the study setting.

The marked imbalance between emergency and elective Caesarean sections suggests that the study population was largely composed of high-acuity cases, which may have implications for postoperative outcomes, including infection risk, length of hospital stay, and maternal morbidity.

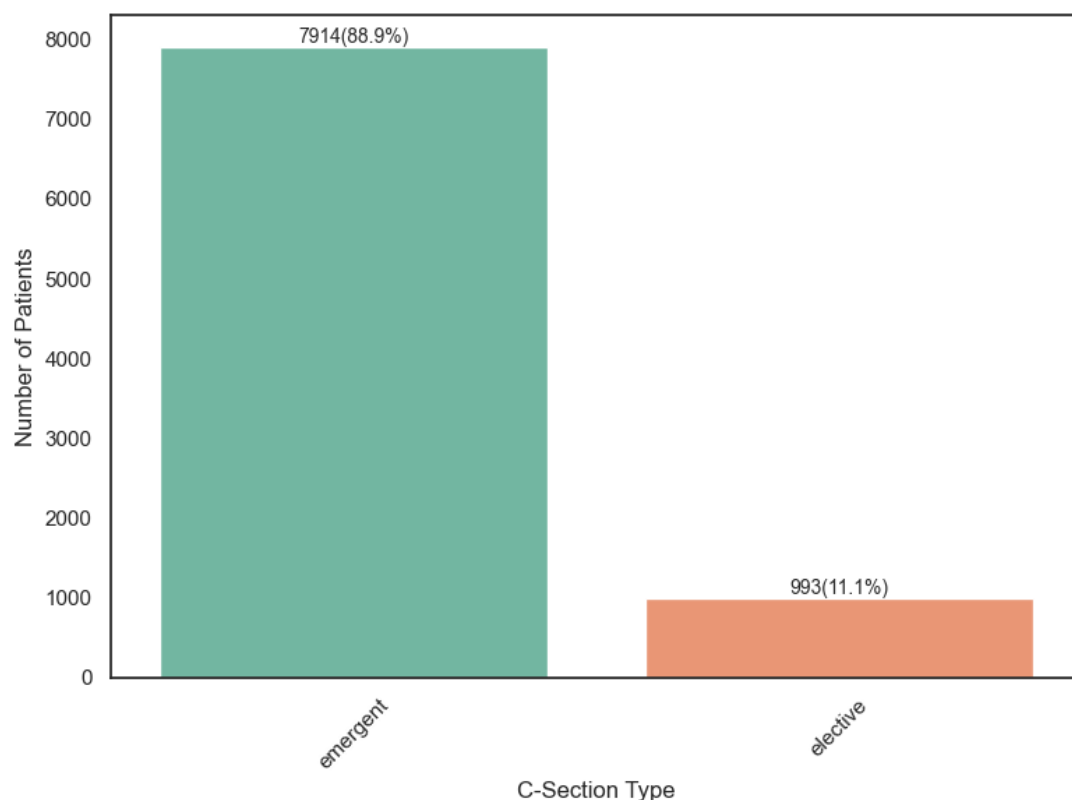


Figure 3.13: Distribution of C-section types showing the number of emergent and elective cases.

3.6.3.2 Distribution of Indications for Caesarean Section

The analysis of indications for Caesarean section demonstrated that previous uterine scar was the most common indication, accounting for the largest proportion of cases by a substantial margin. This was followed by obstructed labor and fetal distress, which were also prominent contributors to surgical delivery.

Moderately represented indications included prolonged labor, pre-eclampsia, big baby (macrosomia), and cephalopelvic disproportion (CPD). Other obstetric conditions such as oligohydramnios, breech presentation, multiple pregnancy, contracted pelvis, and antepartum hemorrhage (APH) were observed less frequently.

Relatively rare indications included cord prolapse, intrauterine fetal death (IUFD), ruptured uterus, premature labor, severe polyhydramnios, and poor maternal effort, each contributing only a small fraction of total cases.

The predominance of previous uterine scar as an indication suggests a high rate of repeat Caesarean sections within the study population. This finding may reflect evolving obstetric practice patterns, referral dynamics, or limited access to vaginal birth after Caesarean (VBAC). The substantial contribution of obstructed labor and fetal distress further highlights the burden of acute intrapartum complications in the study setting.

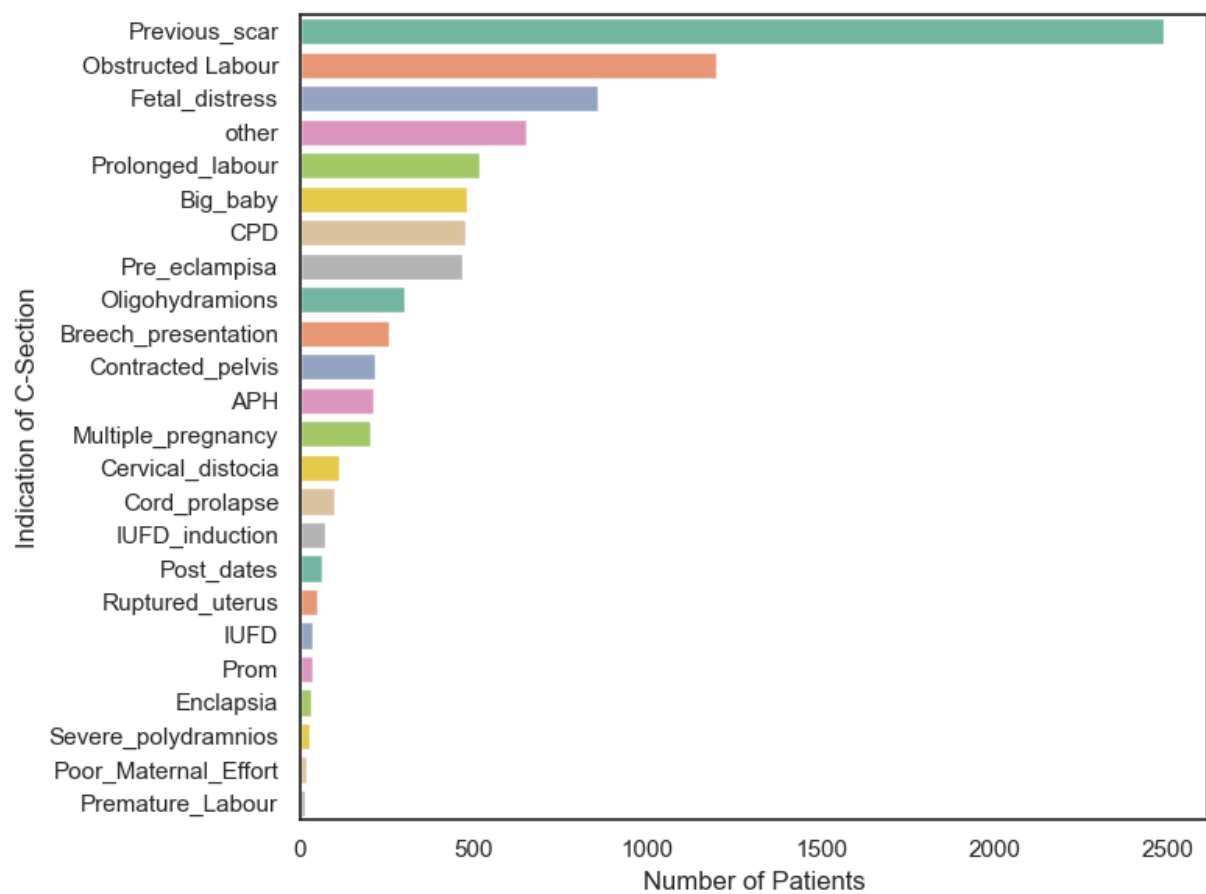


Figure 3.14: Distribution of indications for C-section showing patient counts across multiple clinical causes.

3.6.3.3 Distribution of Caesarean Sections by Surgical Personnel

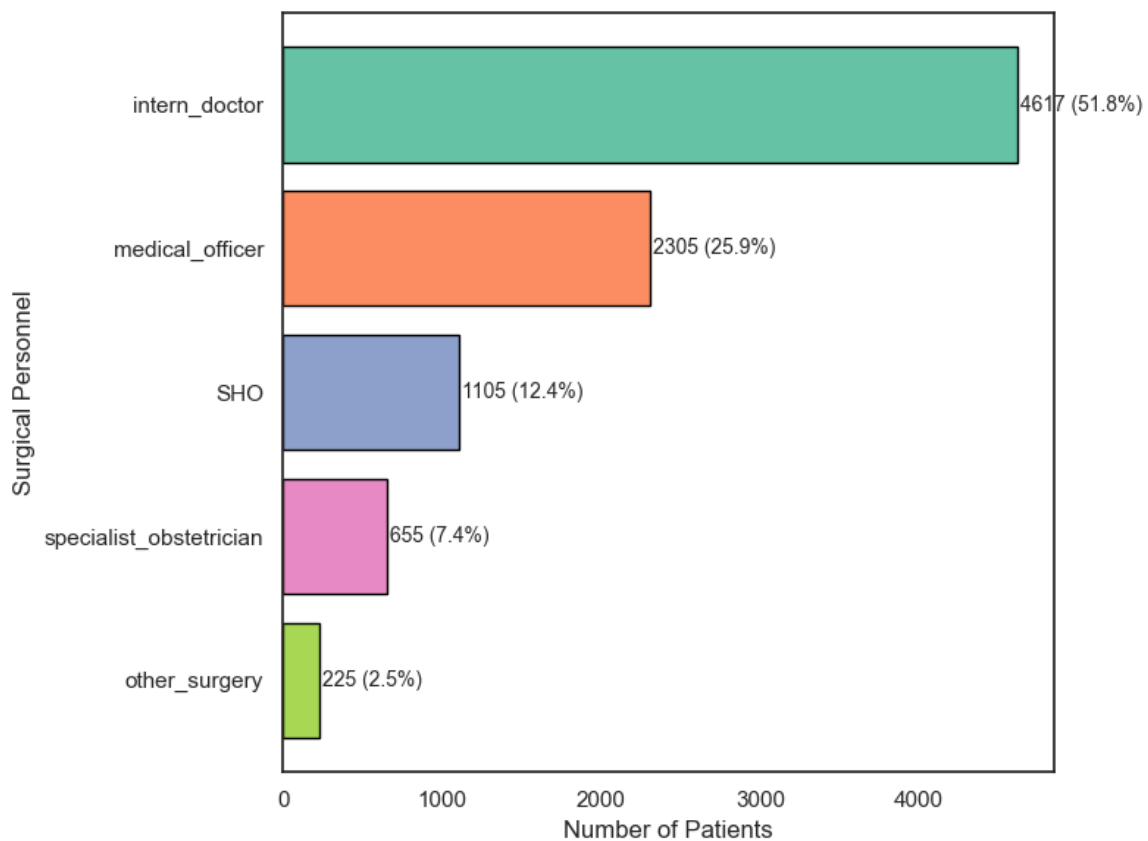


Figure 3.15: Distribution of surgical personnel involved in C-sections, showing the number of patients managed by each cadre.

The analysis demonstrated that the majority of Caesarean sections were performed by intern doctors, accounting for 4617 procedures, representing the largest proportion of surgeries within the study population. This was followed by medical officers, who performed 2305 procedures.

Senior House Officers (SHOs) conducted 1105 Caesarean sections, while specialist obstetricians performed 655 procedures. A small number of cases (225) were categorized under other surgical personnel.

The findings indicate that Caesarean sections in this setting are predominantly carried out by junior medical staff, particularly intern doctors, with comparatively fewer procedures performed by specialist obstetricians. This distribution may reflect task-sharing practices, workforce structure, patient load, and the availability of specialist care within the facility.

3.6.3.4 Distribution of Caesarean Sections by Anesthesia Personnel

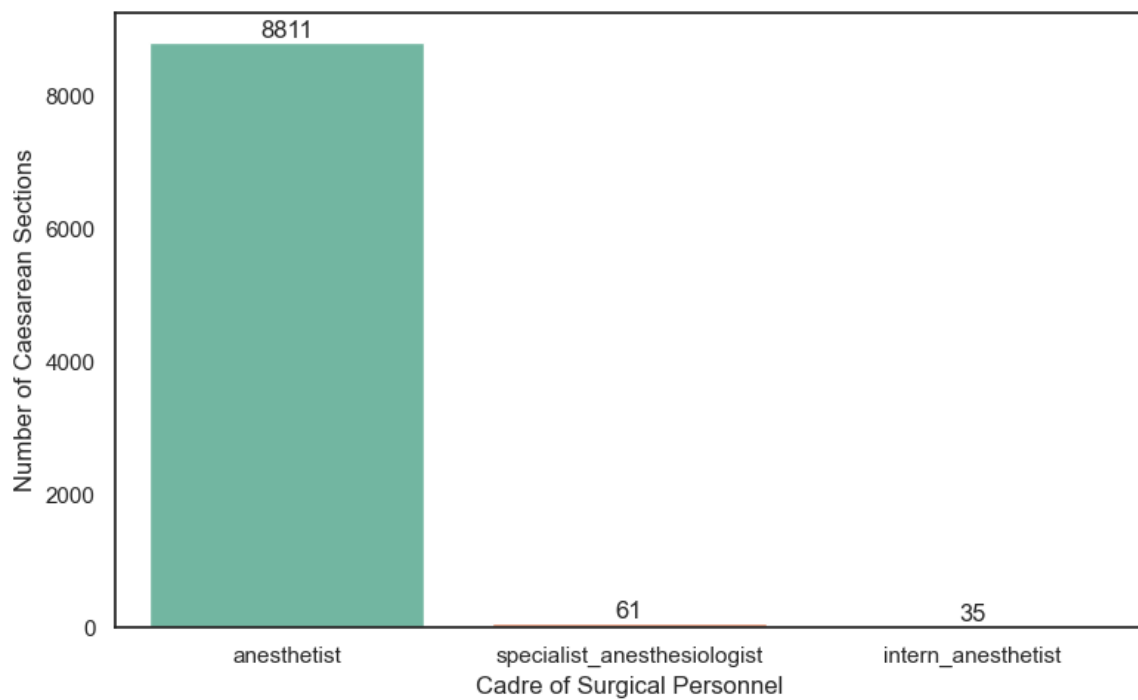


Figure 3.16: Distribution of cadres of surgical personnel involved in Caesarean sections.

The findings reveal a marked imbalance in provider involvement. The overwhelming majority of caesarean sections (8811 cases) were conducted under the care of anesthetists, while specialist anesthesiologists accounted for only 61 cases and intern anesthetists for 35 cases. These results demonstrate a heavy reliance on non-specialist anesthesia providers within the hospital setting.

3.6.3.5 Distribution of HIV Status among Patients

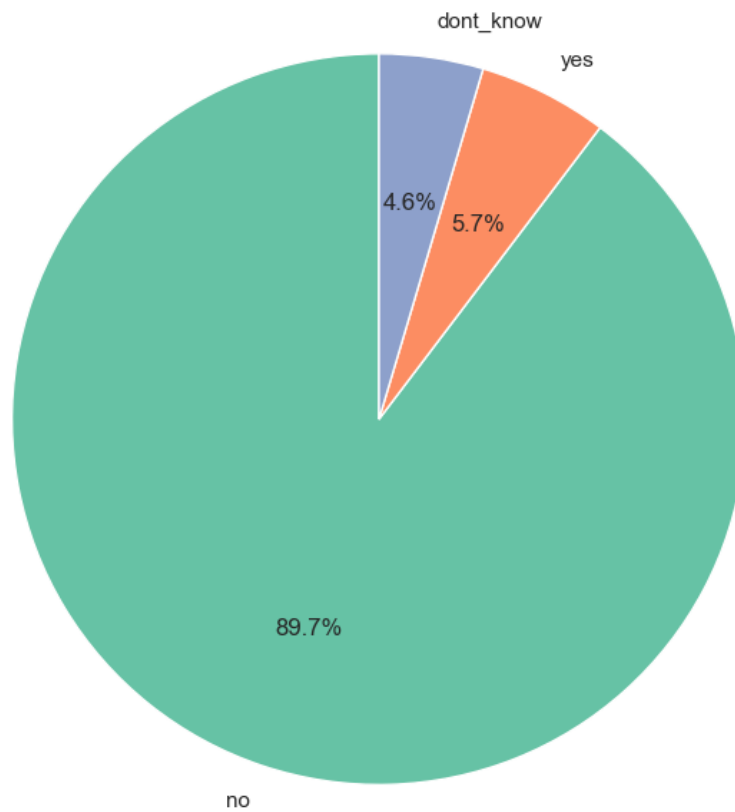


Figure 3.17: Distribution of HIV Status among Patients

The findings demonstrate that the majority of patients were HIV negative, accounting for 7990 cases (89.7%). Patients living with HIV comprised 509 cases (5.7%), while 408 patients (4.6%) had an unknown HIV status at the time of care. The predominance of HIV-negative patients reflects the general population distribution within the study setting.

3.6.3.6 Distribution of Patient Skin Preparation Methods

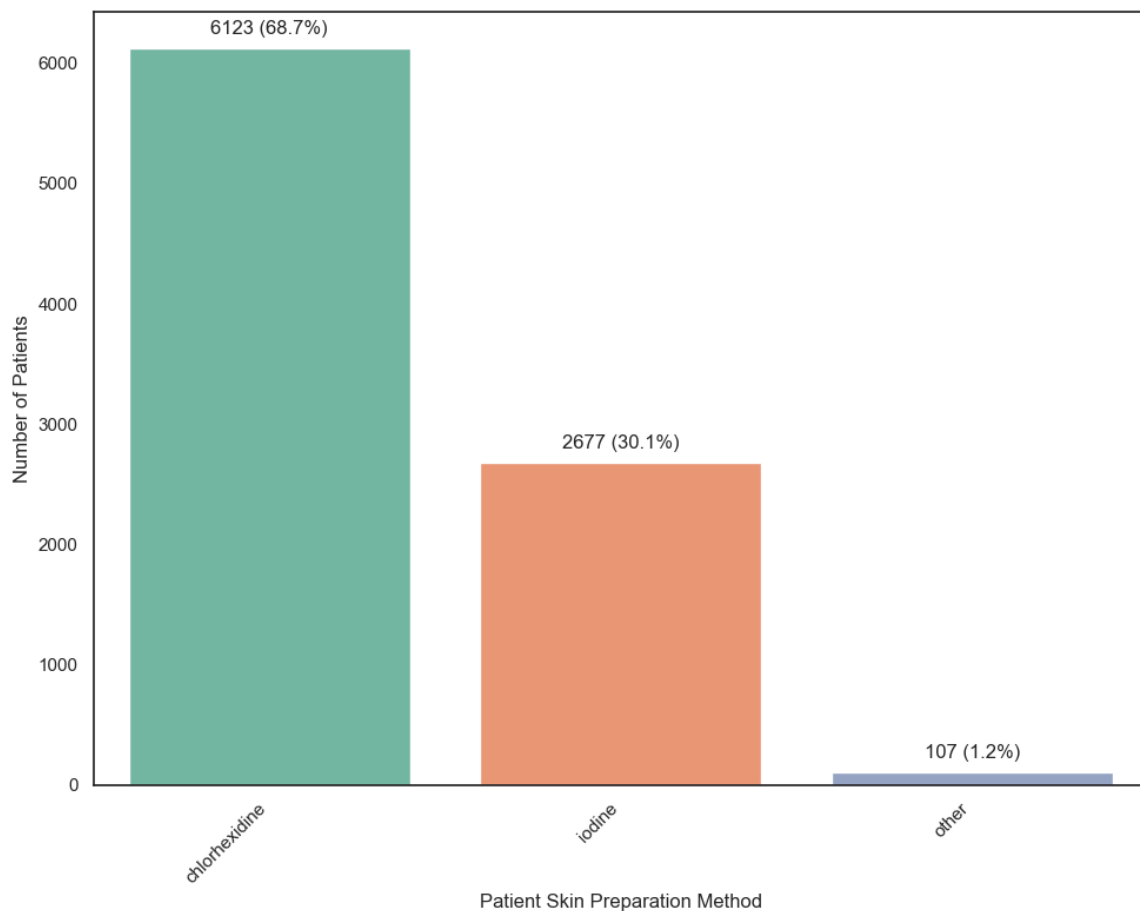


Figure 3.18: Distribution of Patient Skin Preparation Methods

This study evaluated the distribution of patient skin preparation methods used prior to caesarean section. The findings indicated that chlorhexidine was the predominant antiseptic agent, utilized in 6123 cases (68.7%). Iodine-based preparations were used in 2677 cases (30.1%), while alternative methods accounted for only 107 cases (1.2%). The marked preference for chlorhexidine suggests adherence to contemporary infection prevention guidelines.

3.6.4 Data Structure

Data structure focuses on the identification of the size, shape, features, and dimensionality using the Python **pandas** library to support this process. The dataset contained 8,918 records and 34 features. The features are grouped into 3 different data types: 2 float, 2 integer, and 49 object-type features. The `info()` method of the **pandas** DataFrame was used to inspect the dataset structure, including data types and missing values. For example, variables such as `Was_antibiotic_Prophylaxis_given_Pre_Operative`, `Was_the_patient_taught_about_hygiene_and_wound_care`, and `Were_antibiotics_given_post_operative` were examined.

Several of the features were dropped during data pre-processing because they were not important for this analysis. The data was organised in a two-dimensional array.

The dataset consists of 8,918 patient records with 34 features, capturing both demographic and clinical information relevant to surgical outcomes. The features are categorised into different data types: numeric, categorical, and object types representing dates or text entries.

To inspect the dataset structure, including data types and missing values, the `info()` method from the `pandas` library was used. This helped identify features that contained missing values.

The data was structured as a two-dimensional array, suitable for further analysis such as feature selection, encoding categorical variables, and predictive modelling. This organization allows for efficient handling of mixed data types and supports downstream machine learning workflows.

3.7 Analytical Environment

The data analysis and model development were conducted using the Python programming language, which is widely recognised for its versatility in handling and analysing large datasets. Python’s extensive ecosystem of libraries, including Scikit-learn, Pandas, and Numpy, provides robust tools and functions for implementing machine learning algorithms and building intelligent applications.

Scikit-learn was selected due to its optimised implementations of classical machine learning algorithms suitable for structured clinical tabular data. Deep learning frameworks were not prioritised because the dataset is moderate in size and primarily tabular, where tree-based ensemble methods typically outperform neural architectures.

The dataset was obtained in spreadsheet format (`.xlsx`), which allowed for seamless importation and manipulation within the Python environment. All analyses were performed on an HP laptop equipped with a Core i5 processor, 8 GB RAM, 1 TB storage, running Microsoft Windows 11.

Experiments were executed using reproducible pipelines using fixed random seeds.

Component	Tool
Programming	Python 3.11
Notebook	Jupyter
Processing	Pandas, NumPy
Visualization	Matplotlib
Machine Learning	Scikit-learn
Imbalance Handling	Imbalanced-learn
Explainability	SHAP
Boosting	XGBoost

Table 3.3: Software Tools and Libraries Used in the Study

3.7.1 Overview of Analytical Pipeline

This study followed a structured analytical pipeline to assess the feasibility of explainable machine learning models for early detection of postpartum Surgical Site Infections (SSI), and to identify data gaps in routine hospital records that limit predictive performance.

The process began with data extraction and integration, where clinical and follow-up data were obtained from the study database. The raw dataset then underwent data cleaning and pre-processing, which included correcting invalid entries, handling missing values using Multiple Imputation by Chained Equations (MICE), encoding categorical variables, and scaling contin-

uous features to ensure compatibility with machine learning algorithms. In addition, feature engineering was performed to derive and refine variables relevant to SSI risk prediction.

To ensure unbiased model evaluation and prevent data leakage, the dataset was first partitioned into training and testing subsets. This step was performed prior to any resampling procedures, preserving the independence of the test data.

Given the significant class imbalance in SSI outcomes, the Synthetic Minority Over-sampling Technique for Nominal and Continuous variables (SMOTENC) was applied exclusively to the training dataset. This approach enabled the generation of synthetic samples for the minority class while appropriately handling mixed data types, thereby improving the model's ability to learn patterns associated with SSI cases without compromising the validity of model evaluation.

Following this, four machine learning models that is; Logistic Regression, Random Forest, Support Vector Machine (RBF kernel), and XGBoost , were trained and optimised using the balanced training data.

Model performance was evaluated on the untouched test dataset using multiple metrics, including accuracy, ROC-AUC, precision, recall, and F1-score. Particular emphasis was placed on recall and F1-score, given the clinical importance of correctly identifying SSI cases in an imbalanced setting. Additional evaluation tools, including Precision-Recall curves and confusion matrices, were used to provide deeper insight into model behavior under class imbalance.



Figure 3.19: Overview of the analytical pipeline for Surgical Site Infection (SSI) prediction.

3.8 Feature Engineering

To ensure methodological rigor, prevent data leakage, and support robust predictive modelling, a structured feature selection and engineering process was implemented prior to model devel-

opment.

The cleaned dataset comprised 8,875 records, with 33 predictor variables and one binary outcome variable (`final_outcome`). The dataset was first split into training and testing subsets using an 80:20 ratio, resulting in 7,100 records for training and 1,775 records for testing. This separation ensured that model evaluation was performed on unseen data, thereby providing an unbiased estimate of generalization performance.

Feature engineering was then applied consistently to both the training and test datasets. New variables were derived from existing features to better capture clinically meaningful patterns. These included time-based variables such as `length_of_stay` and `Duration_of_surgery_minutes`, as well as composite and categorical risk indicators such as `long_surgery`, `surgical_risk`, `hygiene_score`, and categorized surgical indication variables (`low_risk_indication`, `moderate_risk_indication`, and `high_risk_indication`).

During preprocessing, data quality issues were addressed. Negative or implausible values introduced during feature derivation (e.g., negative values in `Duration_of_surgery_minutes`) were identified and replaced with missing values (NaN). These were subsequently handled using Multiple Imputation by Chained Equations (MICE) via the Iterative Imputer, applied separately to maintain data integrity across the training and test datasets.

Categorical variables were processed based on their characteristics. Low-cardinality variables were encoded using one-hot encoding, while ordinal variables such as `Age_group` were transformed using ordinal encoding to preserve inherent ordering. High-cardinality variables, including `Surgery_by` and `Anesthesia_by`, were also encoded using ordinal encoding after grouping rare categories into an “Other” category (based on a defined frequency threshold). This step reduced sparsity and improved model stability.

To avoid data leakage, variables that directly reflected the outcome or post-outcome clinical observations were explicitly removed from the feature set. These included post-operative wound assessment indicators such as purulent discharge, fever, wound reopening, redness, swelling, and increased wound pain. Additional leakage-prone variables such as patient identifiers, redundant site variables, and derived indicators closely tied to post-outcome information were also excluded from the modelling dataset.

After cleaning, encoding, and feature engineering, the final dataset was prepared by merging imputed numerical features with encoded categorical variables. The resulting feature matrices were aligned between training and test sets to ensure consistent dimensionality. Any remaining missing values were handled, and all features were converted to numerical format to facilitate model training.

To address class imbalance in data, the Synthetic Minority Over-sampling Technique for Nominal and Continuous features (SMOTENC) was applied to the training dataset only. This approach accounts for both categorical and continuous variables when generating synthetic samples, resulting in a balanced training dataset.

Finally, feature scaling was performed using standardization (`StandardScaler`) to ensure that all numerical variables contributed equally during model training, particularly for algorithms sensitive to feature magnitude such as Logistic Regression and Support Vector Machines.

After completing feature selection and engineering, the dataset was fully prepared for modelling. The final analytical pipeline ensured that all predictors used were available prior to or at the time of surgery, were free from leakage, and were appropriately transformed for machine learning algorithms. This resulted in a robust feature set that supported reliable prediction of Surgical Site Infections while maintaining clinical relevance and interpretability.

3.9 Model Development

Model development followed a structured and comparative approach aimed at identifying the most suitable algorithm for predicting Surgical Site Infections (SSI). The process began with the implementation of a baseline model using Logistic Regression. This model was selected due to its simplicity, computational efficiency, and high interpretability, providing a benchmark against which more complex models could be evaluated.

Following the baseline model, several advanced machine learning algorithms were implemented to capture complex and non-linear relationships within the clinical dataset. These included Decision Tree, Random Forest, Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). Decision Trees were utilized for their ability to model non-linear decision boundaries and provide intuitive rule-based interpretations[48]. Random Forest, an ensemble learning method, was employed to improve predictive accuracy and reduce overfitting by aggregating multiple decision trees.

Support Vector Machine (SVM) was applied due to its effectiveness in handling high-dimensional data and its ability to identify optimal decision boundaries. Additionally, XGBoost, a gradient boosting algorithm, was implemented to further enhance model performance through sequential learning and optimization of prediction errors.

To address class imbalance in the dataset, particularly the relatively low occurrence of Surgical Site Infection (SSI) cases, the Synthetic Minority Over-sampling Technique for Nominal and Continuous features (SMOTENC) was applied prior to model training. Unlike standard SMOTE, which assumes that all features are continuous, SMOTENC was specifically designed to handle datasets comprising both numerical and categorical variables by treating categorical features differently during the generation of synthetic samples.

This approach ensured that the synthetic observations preserved the integrity of categorical feature distributions while enhancing the representation of the minority class. Consequently, the models were trained on a more balanced and realistic dataset, which improved their ability to learn meaningful patterns associated with SSI and enhanced predictive performance, particularly in detecting minority class cases.

3.10 Model Performance Evaluation and Validation

The performance of the developed machine learning models was evaluated using a combination of classification metrics to assess both predictive accuracy and clinical relevance in identifying Surgical Site Infections (SSI). Given the imbalanced nature of the dataset, particular emphasis was placed on metrics that effectively capture model performance on the minority class.

Key evaluation metrics included accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). While accuracy provided an overall measure of correctness, it was not solely relied upon due to class imbalance. Greater importance was placed on recall (sensitivity), as it reflects the model's ability to correctly identify patients who develop SSI, thereby minimizing false negatives, which is critical in clinical settings. Precision was also considered to evaluate the proportion of correctly identified positive cases, while the F1-score provided a balanced measure between precision and recall.

A confusion matrix was generated for each model to provide a detailed breakdown of true positives, true negatives, false positives, and false negatives. This enabled a deeper understanding of model behavior, particularly in terms of misclassification patterns and the trade-offs between sensitivity and specificity.

To further assess model discrimination ability, Receiver Operating Characteristic (ROC) curves were plotted and the corresponding ROC-AUC values computed. In addition, Precision-Recall (PR) curves were utilized, as they are more informative for imbalanced datasets by focusing on the performance of the minority class (SSI cases) [49]. These graphical evaluations provided insight into model performance across different classification thresholds.

3.11 Model Interpretability and Insights

To ensure that the developed models are transparent, explainable, and suitable for clinical application, model interpretability techniques were employed. Given the importance of trust and accountability in healthcare settings, these techniques were used to provide both global and local explanations of model predictions.

Feature importance analysis was conducted, particularly for ensemble models such as Random Forest and XGBoost, to identify the most influential variables contributing to the prediction of Surgical Site Infections (SSI). This enabled the identification of key risk factors, such as duration of surgery, length of hospital stay, hygiene-related practices, and patient clinical characteristics, thereby aligning model outputs with clinical knowledge.

In addition, SHAP (Shapley Additive Explanations) values were utilized to provide a more detailed and consistent interpretation of model predictions. SHAP analysis offered both global interpretability highlighting overall feature influence across the dataset and local interpretability, explaining individual predictions for specific patients. This allowed for a clear understanding of how different features increased or decreased the predicted risk of SSI.

These interpretability tools were further supported by visualizations, such as feature importance plots and SHAP summary plots, which enhanced the communication of results to both technical and non-technical stakeholders.

3.12 Alignment Between Objectives and Methods

Objective	Method	Output
Identify predictors	EDA + SHAP	Important variables
Compare models	ML evaluation	Best model
Explain predictions	SHAP	Interpretation

Table 3.4: Summary of Objectives, Methods, and Outputs

3.13 Validation and Clinical Contextualization

Once the model was trained and evaluated, it was validated using a separate test set to confirm its performance on unseen data. The results was contextualized for the Ugandan healthcare environment, considering factors such as diagnostic workflows, resource limitations, and common comorbidities in local populations. This step will ensure the model’s applicability and relevance in the settings where it is intended to be deployed. Clinical feedback may also be sought to interpret findings and assess practical integration opportunities.

Following model training and internal evaluation, final validation was conducted using a held-out test dataset to assess model performance on previously unseen data. This approach ensured

an unbiased evaluation of the model’s generalizability and its ability to maintain predictive performance beyond the training phase.

The validation process focused on confirming the consistency of key performance metrics, including recall, F1-score, and ROC-AUC, particularly for the minority class (Surgical Site Infections). This step was critical in verifying that the model did not over fit the training data and could reliably identify high-risk patients in real-world scenarios.

Beyond statistical validation, the model outputs were interpreted within the context of the Ugandan healthcare system. Consideration was given to local clinical practices, including peri-operative care protocols, variability in infection prevention measures, and resource constraints such as limited diagnostic capacity and workforce challenges.

The inclusion of context-specific engineered features such as hygiene-related practices, categorized surgical indications (low, moderate, and high risk), and duration-based variables enhanced the model’s relevance to the local setting. These features reflect real-world clinical conditions and workflows, thereby improving the practical utility of the model.

3.14 Ethical Considerations

This study utilized secondary data obtained from the Surgical Site Infection (SSI) surveillance component of the Healthcare-Associated Infections (HAI) programme implemented by the Infectious Diseases Institute in collaboration with participating Regional Referral Hospitals. Ethical approval to access and use the surveillance data for research purposes was obtained through the relevant Research Ethics Committee (REC) and institutional authorization processes prior to commencement of analysis.

Because the study relied exclusively on existing surveillance records and did not involve direct interaction with patients, no additional participant recruitment or intervention was conducted. All analyses were performed on de-identified datasets to protect patient privacy and confidentiality. Personally identifiable information, including patient names and identifiers, was removed prior to data processing and analysis.

Access to the dataset was restricted to authorized study personnel only. Data files were stored on password-protected and encrypted devices and handled in accordance with institutional data governance procedures. Data extraction, cleaning, and modelling workflows were implemented using secure and reproducible analytical environments to minimize risks associated with unauthorized access or data manipulation.

The study adhered to ethical principles of respect for persons, beneficence, non-maleficence, and justice. Findings will be reported in aggregate form to ensure that no individual participant can be identified in publications, presentations, or dissemination materials.

Where applicable, this study complied with institutional research guidelines and national requirements governing the use of health data for research purposes.

Chapter 4

Results

OBJECTIVE 1:

4.1 Descriptive statistics

The mean age of the participants was 26.3 years ($SD = 6.0$), indicating that most mothers were within the younger reproductive age group a pattern consistent with national maternity demographics in Uganda.

With regard to HIV status, the majority of patients were HIV-negative, accounting for 7,990 cases (89.7%). Patients living with HIV constituted 509 cases (5.7%), while 408 mothers (4.6%) had an unknown HIV status at the time of care. This distribution aligns with the general HIV prevalence patterns among reproductive-age women in similar settings.

The preoperative physical status of patients, assessed using the American Society of Anesthesiologists (ASA) classification, showed a mean ASA score of 1.2 ($SD = 0.46$). This suggests that the majority of women were classified as ASA I, indicating they had no significant systemic illness prior to surgery.

The dataset revealed a strong predominance of emergency Caesarean sections, which accounted for 7,942 procedures (88.9%). In contrast, elective Caesareans constituted 992 cases (11.1%). This distribution reflects a high burden of acute obstetric complications and limited opportunities for planned surgical deliveries, a common characteristic of referral hospitals serving high risk populations.

The majority of Caesarean sections were performed by intern doctors, who handled 4,617 procedures (51.7%). This was followed by: Medical Officers: 2,305 cases (25.8%), Senior House Officers (SHOs): 1,105 cases (12.4%), Specialist Obstetricians: 655 cases (7.3%), other providers: 252 cases (2.8%). This distribution highlights a workforce structure heavily reliant on junior medical staff, likely due to staffing constraints and high patient volumes in the study setting.

A similar pattern was observed in anesthesia provision. Anesthetists conducted anesthesia for 8,811 procedures (98.6%), while: Specialist anesthesiologists handled only 61 cases (0.7%), and Intern anesthetists performed 35 cases (0.4%). This demonstrates substantial reliance on non-specialist anesthesia providers, which is typical in resource limited health systems.

Chlorhexidine was the most commonly used antiseptic agent, applied in 6,123 procedures (68.7%). Iodine based skin preparation was used in 2,677 cases (30.1%), while other meth-

ods accounted for 107 cases (1.2%). The dominant use of chlorhexidine suggests adherence to infection prevention recommendations favoring its superior antimicrobial effectiveness.

The incidence of postpartum Surgical Site Infection (SSI) in the study population was 8.1% ($n = 724$), while the majority of patients did not develop SSI, accounting for 91.9% ($n = 8,210$).

4.1.1 Bivariate Analysis of the numerical variables with the Target Variable

A bivariate analysis was conducted to examine the relationship between selected continuous variables and the binary outcome variable *final_outcome*. Boxplots were employed to compare the distribution, central tendency, and dispersion of each numerical variable across infection status categories (“1” and “0”).

The results of the bivariate analysis revealed substantial overlap between the infected and non-infected groups for most continuous variables. Although slight differences were observed in antibiotic duration and length of hospital stay—both of which are plausibly related to infection management—the overall distributions indicate that no single numerical variable exhibits a strong unadjusted association with infection status.

These findings suggest that infection risk is likely influenced by a combination of factors rather than any isolated continuous predictor, highlighting the need for multivariable modelling to better understand the determinants of postoperative infection.

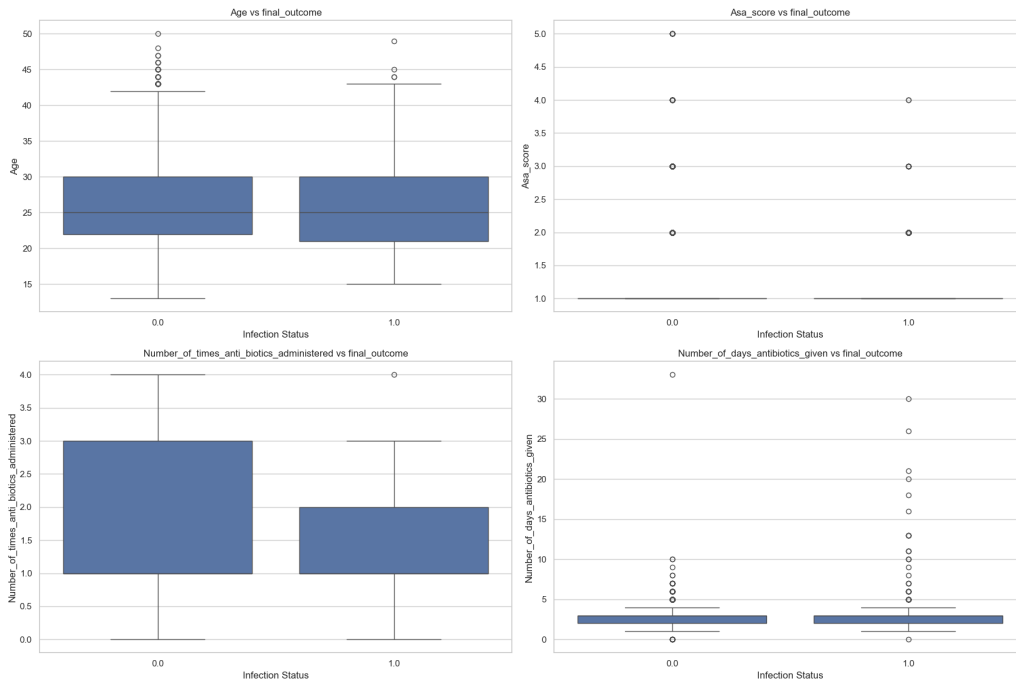


Figure 4.1: Boxplots illustrating the distribution of key continuous variables stratified by surgical site infection (SSI) outcome.

4.1.2 Examining the Target Variable

The class distribution of the outcome variable (*final_outcome*) indicates a pronounced imbalance between patients diagnosed with surgical site infection (SSI) and those without infection. A total of 8,166 patients (91.9%) did not develop SSI (No SSI – 0), while 719 patients (8.1%)

were diagnosed with SSI (SSI = 1). This demonstrates that SSI cases represent a small minority of the overall sample, resulting in a highly imbalanced dataset in which the vast majority of patients fall into the non-SSI category. Such imbalance is common in clinical datasets[46], where adverse postoperative events occur relatively infrequently.

The observed SSI prevalence of 8.1% indicates that postoperative infection was relatively uncommon within the study population. However, the substantial class imbalance underscores the need for careful methodological considerations during predictive modelling to ensure that infection cases are accurately identified and not overshadowed by the majority class.

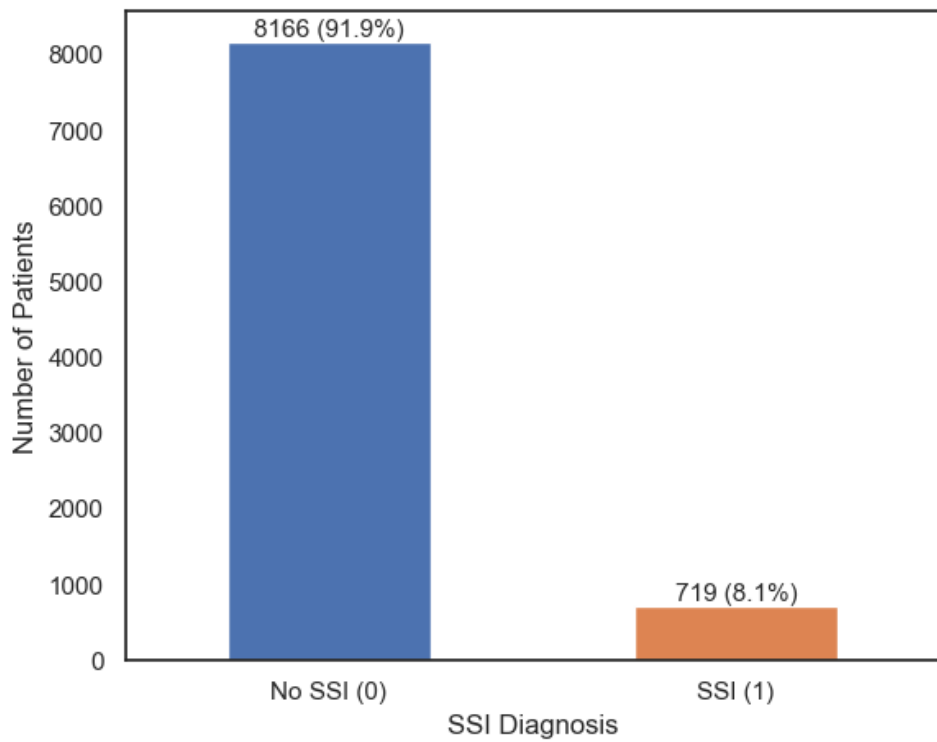


Figure 4.2: Distribution of Surgical Site Infection (SSI) Diagnosis Among Patients

OBJECTIVE 2:

4.2 Model design and Performance comparison

Table 4.1: Performance Evaluation of the Models

Rank	Model	Accuracy	ROC-AUC	F1 Score	Precision	Recall (Sensitivity)
1	Logistic Regression	0.1031	0.6528	0.1523	0.0825	0.9931
2	Random Forest	0.2992	0.6064	0.1640	0.0908	0.8472
3	XGBoost	0.7932	0.5884	0.1716	0.1271	0.2639
4	SVM (RBF kernel)	0.5403	0.5798	0.1774	0.1038	0.6111

4.2.1 Model Development

To develop predictive models for postpartum surgical site infections (SSI), a structured machine learning pipeline was implemented, incorporating data preprocessing, class imbalance handling,

model training, and performance evaluation. The dataset was divided into training and testing sets to enable unbiased evaluation of model performance. Typically, a stratified split was applied 80:20 ratio to ensure that the minority class (SSI cases) was proportionally represented in both subsets, given the observed class imbalance (8.1% SSI vs 91.9% non-SSI).

Given the highly imbalanced nature of the outcome variable, Synthetic Minority Over-sampling Technique Synthetic Minority Over sampling Technique for Nominal and Continuous (SMOTENC) features was applied to the training dataset to improve the representation of the minority class. The test dataset was left unchanged to reflect real-world data distribution.

Four machine learning algorithms were selected based on their complementary strengths in classification tasks and interpretability considerations: Logistic Regression was included as a baseline model due to its simplicity, interpretability, and widespread use in clinical research[50]. Random Forest was chosen for its ability to capture non-linear relationships and interactions between variables while providing measures of feature importance. Extreme Gradient Boosting (XGBoost) was selected for its strong predictive performance and robustness in handling structured data. Support Vector Machine (SVM) with RBF kernel was included to model complex decision boundaries in high-dimensional feature space.

These models represent a balance between interpretability (Logistic Regression), ensemble learning (Random Forest, XGBoost), and non-linear classification (SVM).

Given the clinical importance of minimizing missed SSI cases, classification thresholds were adjusted below the default 0.5 probability cutoff. Lower thresholds were explored to improve sensitivity (recall), ensuring that most true SSI cases were correctly identified. This adjustment reflects a clinical preference for early detection, even at the cost of increased false positives.

4.2.2 Model Performance Evaluation

Model performance was evaluated on the independent test dataset using multiple metrics, including accuracy, ROC-AUC, precision, recall (sensitivity), and F1-score. Particular emphasis was placed on recall and F1-score due to the imbalanced nature of the data and the clinical importance of correctly identifying SSI cases.

Accuracy, while commonly reported, was considered insufficient as a standalone metric in this context because it can be misleading in imbalanced datasets. A model predicting all cases as non-SSI would achieve high accuracy but fail to identify true infection cases.

ROC-AUC was included to assess the model’s ability to discriminate between SSI and non-SSI cases across all classification thresholds. However, given the class imbalance, ROC-AUC may overestimate model performance and was therefore interpreted alongside other metrics.

Precision measures the proportion of predicted SSI cases that are truly positive, reflecting the model’s ability to minimize false positives. Recall (sensitivity), on the other hand, measures the proportion of actual SSI cases correctly identified by the model.

In this study, recall was prioritised due to its clinical importance failing to detect an SSI case could lead to delayed treatment and adverse maternal outcomes. The F1-score, which represents the harmonic mean of precision and recall, was used to provide a balanced assessment of model performance, particularly under class imbalance conditions.

4.2.3 Model Performance Results

The performance of the evaluated models varied considerably across the selected metrics. Logistic Regression achieved the highest recall (0.9931), indicating that it correctly identified nearly all SSI cases. However, this came at the expense of very low precision (0.0825) and low accuracy (0.1031), suggesting a high number of false positive predictions. Despite this limitation, the model demonstrates strong potential for clinical screening applications where sensitivity is prioritised.

Random Forest also demonstrated high recall (0.8472) with slightly improved precision (0.0908) and F1-score (0.1640) compared to Logistic Regression. However, its overall discriminative ability, as indicated by ROC-AUC (0.6064), remained moderate.

XGBoost achieved the highest accuracy (0.7932) among all models, but this was accompanied by substantially lower recall (0.2639), indicating poor detection of SSI cases. This suggests that the model was biased toward predicting the majority class (non-SSI), limiting its usefulness in clinical settings where detecting infections is critical.

The Support Vector Machine (SVM) model provided a more balanced performance, with moderate recall (0.6111) and the highest F1-score (0.1774) among the evaluated models. However, its ROC-AUC (0.5798) indicates limited discriminative performance overall.

4.2.4 ROC Curves for all the models

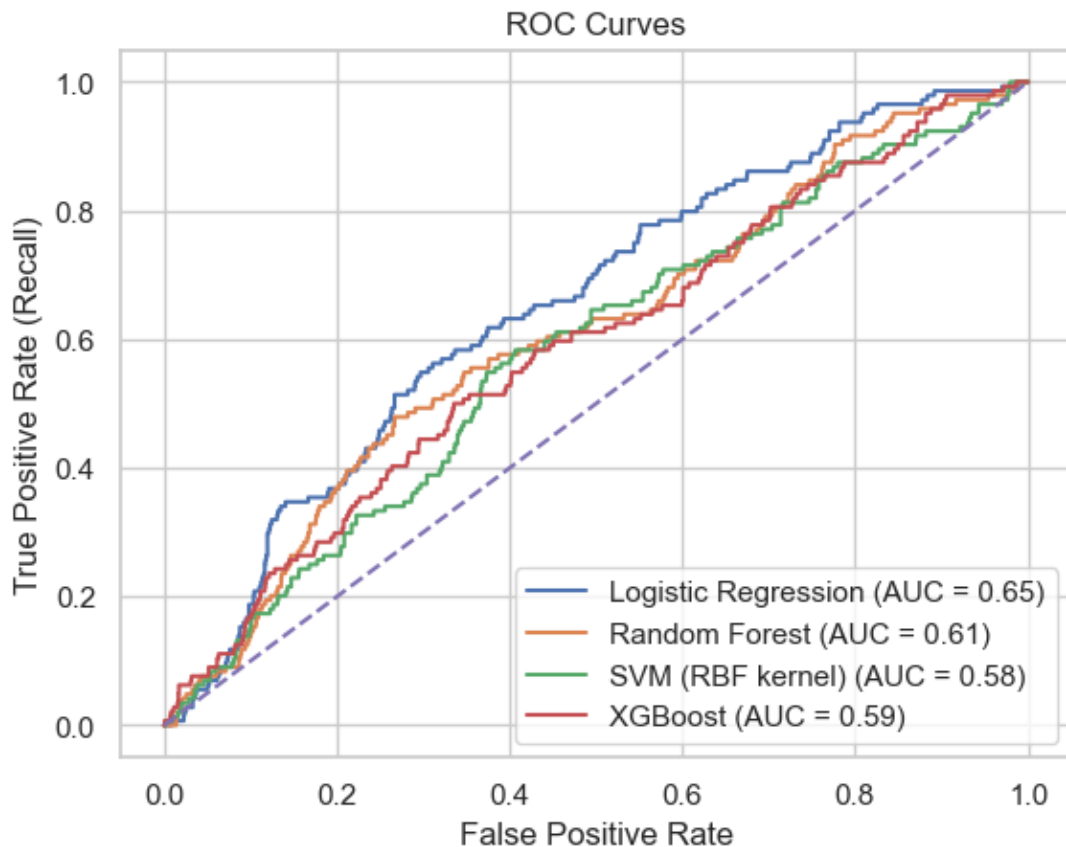


Figure 4.3: ROC Curves comparing Logistic Regression, Random Forest, SVM (RBF kernel), and XGBoost models.

The Receiver Operating Characteristic (ROC) curves provide a visual assessment of each model's ability to discriminate between patients who developed Surgical Site Infections (SSI)

and those who did not across a full range of classification thresholds. The diagonal dashed line represents a no skill classifier ($AUC = 0.50$), serving as a baseline for comparison. Models whose curves rise further toward the top left corner demonstrate stronger discriminative power, as they achieve higher True Positive Rates while maintaining lower False Positive Rates.

From the graph, Logistic Regression exhibits the best overall performance, achieving an AUC of approximately 0.65. Its ROC curve consistently lies above those of the other models across most threshold levels, indicating a superior ability to separate SSI from non SSI cases. Although modest, this performance reflects comparatively stronger discrimination and suggests that Logistic Regression provides a better balance between sensitivity and specificity than the other evaluated models.

The Random Forest model follows with an AUC of about 0.61, indicating moderate discriminative capability. Its curve remains above the no-skill line and reasonably close to that of Logistic Regression, suggesting that Random Forest captures some meaningful structure in the data but is less effective at distinguishing between the two classes.

In contrast, XGBoost ($AUC \approx 0.59$) and SVM with an RBF kernel ($AUC \approx 0.58$) showed weaker performance. Their ROC curves lay closer to the diagonal baseline, indicating only slight improvement over random guessing. This reflects a limited ability to correctly rank patients by risk of SSI, likely influenced by the severe class imbalance in the dataset and the difficulty these models faced in identifying minority-class patterns.

4.2.5 Precision –Recall Curve Analysis for all the models

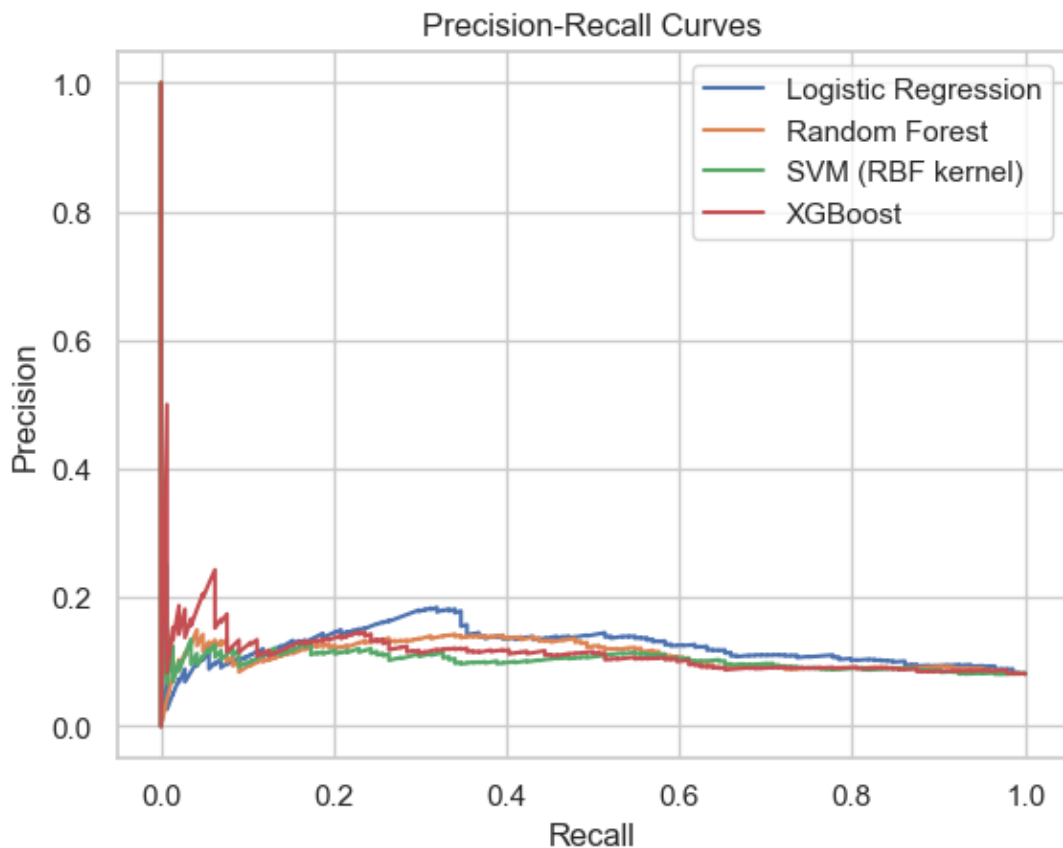


Figure 4.4: Precision–Recall curves for Logistic Regression, Random Forest, SVM (RBF kernel), and XGBoost models.

The Precision–Recall (PR) curves offer additional insight into model performance under the severe class imbalance present in the dataset, where Surgical Site Infections (SSI) represent a small minority of cases. Unlike ROC curves, which can appear overly optimistic in imbalanced settings, PR curves more clearly highlight the difficulty the models face in achieving both high precision and high recall.

As shown in the figure, all models exhibit consistently low precision across most recall levels, indicating that a large proportion of the positive predictions are false positives. Logistic Regression shows a slightly higher and more stable precision curve compared to the other models, suggesting a marginally better balance between identifying true SSI cases and limiting false alarms.

Random Forest and SVM (RBF kernel) display similar behavior, with modest precision values that remain close to those of Logistic Regression but without significant separation. XGBoost shows highly unstable precision at very low recall values, briefly spiking before rapidly declining, reflecting its tendency to identify very few positive cases and perform poorly when it attempts to do so.

Overall, the PR curves highlight the inherent difficulty of the SSI prediction task: while some models achieve reasonable recall, all struggle to maintain precision, reinforcing the challenge of accurately detecting rare clinical outcomes without generating substantial false positives.

4.2.6 Confusion Matrix Analysis

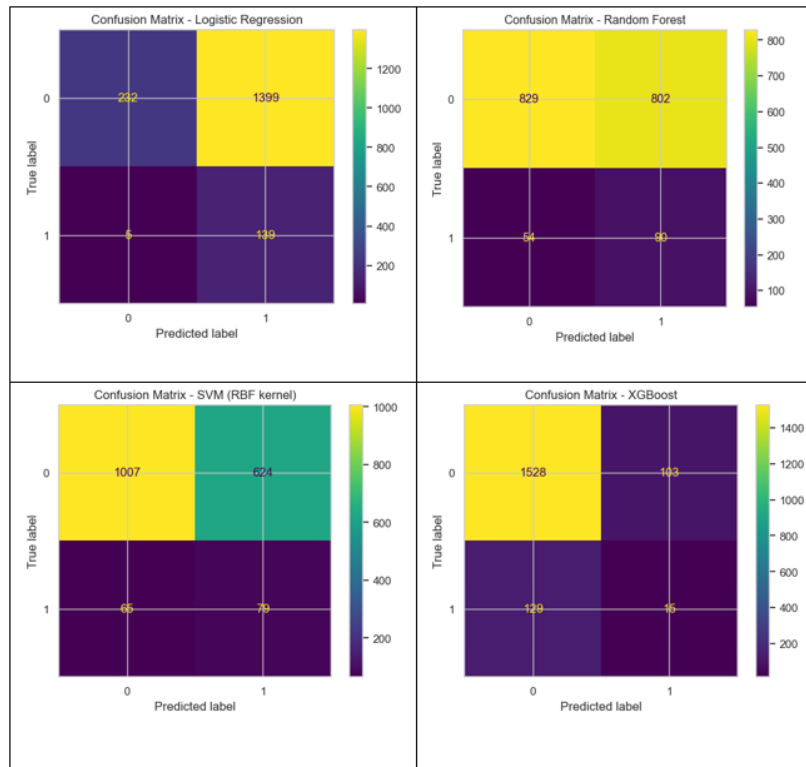


Figure 4.5: Confusion matrices illustrating the classification performance of Logistic Regression, Random Forest, Support Vector Machine (RBF kernel), and XGBoost models.

The confusion matrices for the four evaluated models Logistic Regression, Random Forest, SVM (RBF kernel), and XGBoost provide detailed insight into how each algorithm performs in distinguishing between patients with and without Surgical Site Infections (SSI). Because SSI represents a minority class, the patterns observed in these matrices are critical for understanding each model's strengths and limitations in a clinically imbalanced setting.

Across all models, a common trend emerges: correctly identifying SSI cases (true positives) is significantly more challenging than identifying non SSI cases (true negatives). Logistic Regres-

sion demonstrates the highest sensitivity, correctly identifying almost all infection cases (139 true positives with only 1 false negative).

However, this comes at the cost of an exceptionally high number of false positives (1,399), indicating that the model predicts the positive class excessively. This explains the model’s very high recall but extremely low precision and highlights its tendency to flag many non SSI patients as at risk.

The Random Forest model presents a more balanced but less sensitive profile, capturing 83 true positives while missing 64 actual SSI cases. Although it reduces false positives relative to Logistic Regression, it still produces a substantial number of misclassifications in both directions (802 false positives). This reflects moderate recall and poor precision, consistent with the patterns observed in the ROC and PR curves.

The SVM (RBF kernel) model shows slightly stronger specificity (fewer false positives, 624) but still struggles to detect positive cases effectively, with 69 false negatives and 78 true positives. While it achieves a somewhat improved balance between sensitivity and specificity compared to Logistic Regression and Random Forest, it still fails to identify a considerable number of actual SSI cases, limiting its clinical utility.

In contrast to the other models, XGBoost heavily favors predicting the majority class. It produces very few false positives (133) but misses almost all true infection cases, with 129 false negatives and only 18 true positives. This behavior explains its deceptively high accuracy and poor recall, confirming that the model largely defaults to predicting patients as non SSI. Such performance is unsuitable for clinical early warning systems where failing to detect true infections can have serious consequences.

4.2.7 Feature Importance Results

To better understand the variables contributing most strongly to the Random Forest model’s predictions of Surgical Site Infection (SSI), feature importance scores were extracted and visualized. Random Forest assigns importance based on the extent to which each feature reduces impurity across all decision trees in the ensemble. Higher importance values indicate stronger influence on the model’s decision-making process.

The ranking of the top 20 features (Figure 23) shows that the model relies heavily on variables related to surgical complexity, operative duration, perioperative infection prevention practices, and maternal or procedure related risk factors. These results highlight that both clinical workflow factors and patient-related behaviors play key roles in SSI risk prediction. The prominence of time based features and perioperative interventions aligns with existing evidence showing that prolonged procedures, inadequate prophylaxis, and poor hygiene conditions are major contributors to postoperative infections.

Although Random Forest was not the highest performing model in terms of overall predictive accuracy or ROC AUC, it was deliberately selected as the primary model for interpretability due to its unique strengths in generating reliable and clinically meaningful feature importance insights[47] .

Unlike models such as SVM or XGBoost, which operate through complex, opaque decision boundaries, Random Forest produces transparent, stable importance scores that are easy to interpret and grounded in its ensemble based structure. These importance scores directly reflect how much each feature contributes to reducing prediction error across hundreds of decision trees, offering a robust and intuitive measure of variable influence.

Figure 4.6 highlights the variables that contribute most to the model’s decision-making process, providing insight into key clinical and procedural factors associated with SSI risk.

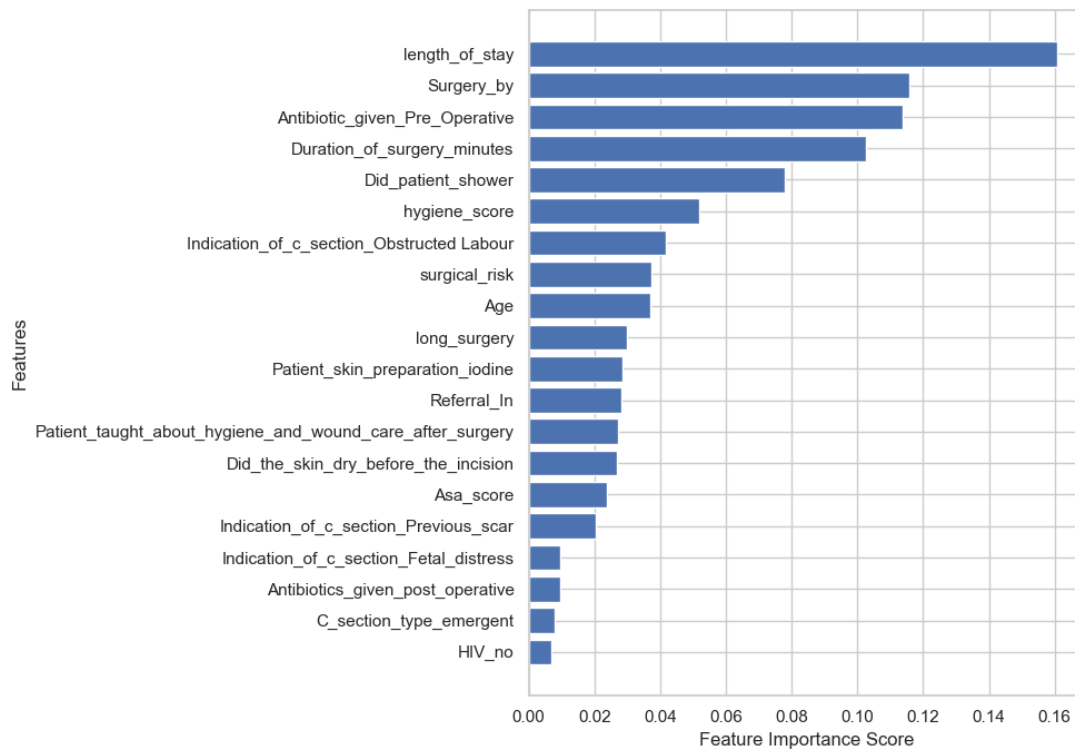


Figure 4.6: Top 20 most influential features for predicting Surgical Site Infections (SSI).

OBJECTIVE 3:

4.3 Explainability using Shapley Additive exPlanations (SHAP)

SHAP was used to determine the contributions of each prediction variable to the model’s predictions. SHAP assigns an importance value to each feature used in the development of the model and also aids in fine-tuning the performance of the model by removing features or variables with low importance values.

4.3.1 SHAP bee swarm plot summary

The SHAP bee swarm plot offers a granular view of how each predictor influences the model’s estimated probability of Surgical Site Infection (SSI). Unlike the feature importance bar chart, which provides only a global ranking of variables, the bee swarm plot displays three essential dimensions of interpretability: (1) the magnitude of each feature’s impact on the model output (distance from zero along the x axis), (2) the direction of its influence (whether it shifts predictions toward higher or lower SSI risk), and (3) the distribution and variability of these effects across individual patients.

Each point represents a single observation, with its position on the x axis indicating how much that feature increases (positive SHAP value) or decreases (negative SHAP value) the predicted likelihood of SSI. The colour gradient encodes feature values, where red indicates higher values

and blue indicates lower values, enabling interpretation of how specific value levels correspond to risk

The plot shows that preoperative bathing and proper skin antiseptic preparation consistently pull predictions toward lower SSI probability, underscoring the protective role of standardized infection prevention measures. In contrast, emergency obstetric indications such as obstructed labor and referral related factors shift predictions toward higher risk, reflecting the clinical reality that prolonged or complicated labor increases infection vulnerability[48].

Operative factors such as long surgery duration also display a clear positive association with SSI likelihood. Interestingly, prolonged antibiotic use appears associated with higher predicted SSI risk; this pattern likely reflects reverse causality, where extended antibiotic administration occurs as a consequence of developing an infection rather than as a cause. These patterns emphasize the importance of interpreting machine learning outputs within their clinical context and distinguishing predictive associations from causal relationships.

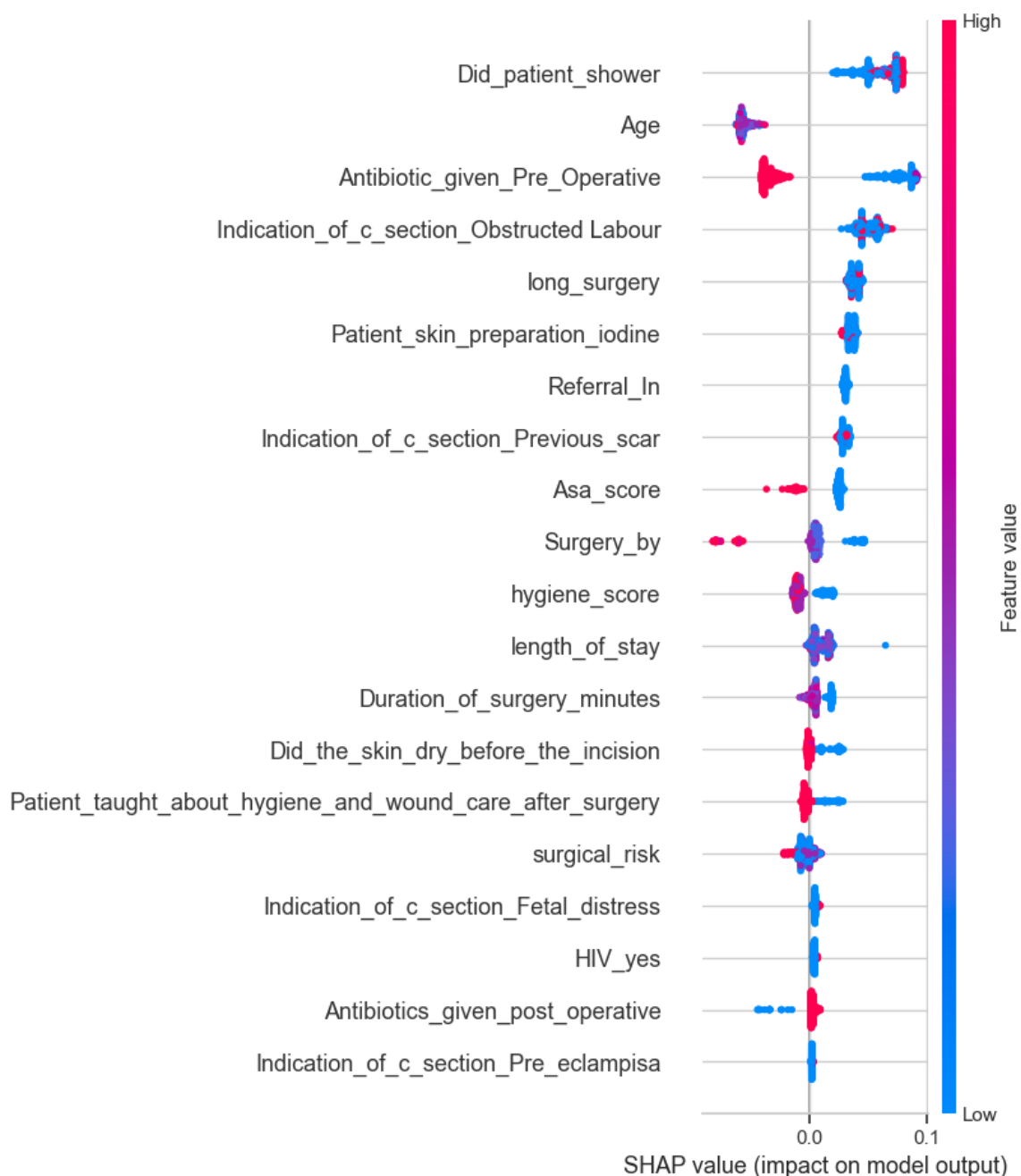


Figure 4.7: SHAP Summary Plot Showing Feature Impact on Model Output

4.3.2 SHAP Waterfall Plot for Individual SSI Risk Explanation

To complement the global interpretability provided by the SHAP bee swarm plot, a SHAP waterfall plot was generated to explain the model's prediction for a single patient. The waterfall plot decomposes the predicted probability of Surgical Site Infection (SSI) into additive contributions from individual features, starting from the model's baseline prediction (the average predicted risk across the entire dataset) and progressively incorporating each patient specific characteristic.

The baseline value for the model was approximately 0.50, representing the average SSI probability before accounting for any individual patient features. Each bar in the plot shows how a specific predictor shifts the model's output either upward (increasing SSI risk) or downward (reducing SSI risk).

Features shown in red contribute positively toward higher risk, while features in blue decrease the predicted risk. The slope of these contributions ultimately leads to the patient's final predicted probability of 0.718, meaning that the model considers this patient to be at substantially higher than average risk of SSI.

Several features exerted strong upward pressure on the prediction. Notably, factors such as preoperative shower status, certain obstetric indications, and perioperative preparation variables collectively increased the patient's estimated risk. Clinical risk scores (e.g., ASA score) also contributed positively, reflecting underlying health status that elevates infection susceptibility.

Conversely, younger age and appropriate administration of preoperative antibiotics reduced the predicted risk, illustrating protective effects that align with established clinical expectations. A small cumulative contribution from numerous less influential variables further adjusted the prediction.

This individualized explanation highlights how the model integrates diverse clinical, demographic, and perioperative factors to arrive at its final prediction. Importantly, the waterfall plot emphasizes that machine learning predictions arise from complex interactions between features, and that not all feature effects are necessarily causal.

The association between certain preoperative practices and increased predicted SSI risk may reflect reverse causality or confounding, where high risk patients receive more intensive perioperative care. Such nuances underscore the need to interpret SHAP outputs in clinical context rather than assuming direct causation.

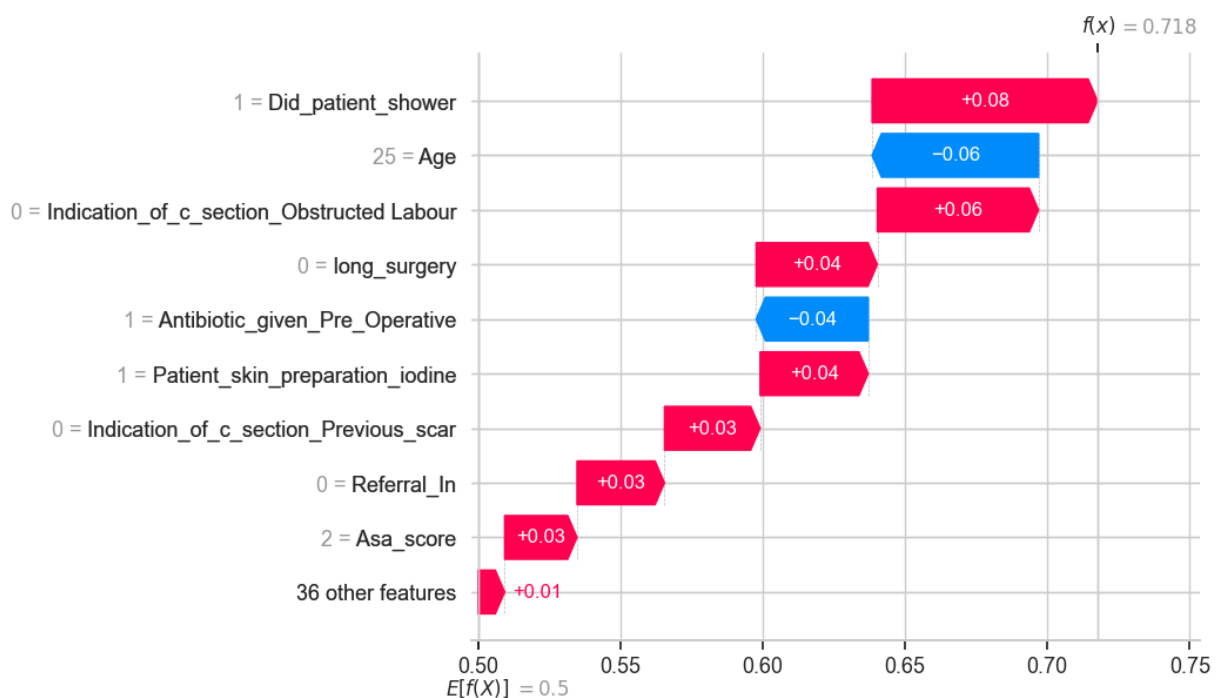


Figure 4.8: SHAP Decision Plot Illustrating Feature Contributions Toward the Predicted Output.

Chapter 5

Discussion

This study evaluated the feasibility of explainable machine learning (ML) models for the early detection of postpartum surgical site infections (SSIs) in Ugandan hospitals. While the study successfully identified key clinical predictors, the results highlight the significant challenges of modelling rare clinical events within routine hospital data.

5.1 Interpretation of Key Findings

The performance of the models was constrained by two primary factors: the 91.9% / 8.1% class imbalance and a weak predictive signal in the available feature set. While Logistic Regression achieved an impressive Recall (0.9931), its Precision (0.0825) was exceptionally low. This suggests that while the model "captured" almost all SSI cases, it did so by casting an extremely wide net.

The fact that complex, non-linear models like XGBoost and Random Forest did not significantly outperform the simpler Logistic Model implies that the "signal" for SSI in routine Ugandan hospital data is not only sparse but also potentially linear. The features used, such as surgery duration and hygiene indicators, act as general risk markers rather than specific diagnostic predictors. Consequently, the models could identify a "high-risk environment," but lacked the granular data (e.g., intraoperative contamination or specific biomarkers) to precisely distinguish an infected patient from a non-infected one who shared similar routine characteristics.

5.2 Comparison with Literature

These results resonate with broader evidence in clinical informatics [49] regarding the difficulty of predicting "rare-event" binary outcomes. Most literature suggests that when the predictive signal is weak, models struggle to achieve high precision. This was evident in this study, where the strongest F1-score was only 0.1774 (SVM), suggesting that while we can identify who might get an infection, we cannot yet pinpoint exactly who will get one without a high rate of false alarms.

Contribution: This study contributes to the literature by demonstrating the viability of Explainable AI (XAI) in an East African clinical context. While many studies focus on high-income settings with granular electronic health records, this research proves that routine, secondary data in Ugandan hospitals, when processed through SHAP explainability, can reveal high-impact predictors like preoperative hygiene and obstructed labour. This moves the conversation from "black-box" accuracy to "glass-box" clinical insight.

5.3 Implication for Practice - Clinical Practice, Policy, Machine Learning Modelling.

Clinical Practice: The models should be implemented as triage-support tools for patient screening rather than diagnostic certainties. Given the high recall of Logistic Regression, clinicians can use the model to flag the "High Risk" 20–30% of the ward. Even with low precision, this allows midwives to prioritise limited bedside monitoring for those most likely to be among the 8.1% of true SSI cases.

Policy: The modest predictive power (ROC-AUC of 0.65) suggests a policy need for standardized digital data collection. To improve AI performance in Uganda, Ministry of Health guidelines should emphasize the recording and completeness of all demographic data ie height and weight, intraoperative antibiotic timing and wound classification, which are currently missing from routine records.

ML modelling: For future maternal health research in LMICs, "Accuracy" should be abandoned as a primary metric. Researchers should instead prioritise Precision-Recall (PR) curves and Cost-Sensitive Learning, which explicitly weight the 8.1% minority class as more "important" than the majority.

5.4 Methodological Reflections

The methodology was defined by the challenge of the 91.9/8.1 class imbalance. While SMO-TENC oversampling helped the models "see" the minority class, it did not fully resolve the overlap between infected and non-infected feature spaces.

The Threshold Sensitivity Analysis revealed a stark reality: while lowering the decision threshold increased recall, it caused a consistent and significant decline in precision, specificity, and accuracy. This confirms that there is no "mathematical shortcut" to fix a weak predictive signal.

In this study, the threshold adjustment was not just a statistical exercise, but a clinical choice: we prioritised the identification of the 8.1% (Sensitivity) at the calculated expense of increased false positives, acknowledging that in a Ugandan context, the cost of a missed infection far outweighs the cost of an extra clinical review.

5.5 Study Limitation

Several limitations of this study should be acknowledged. The limited number of positive cases may have constrained the model's ability to learn robust patterns and generalise to new patients. This is especially relevant for complex algorithms, which typically require larger datasets to achieve optimal performance.

The study relied on secondary data, which limited control over data collection and feature availability. Several potentially important predictors, such as intraoperative contamination, antibiotic prophylaxis, wound classification, and detailed comorbidities, were not available. Missing or unrecorded information may have restricted the discriminatory power of the models.

Despite these limitations, the study provides valuable insights into the feasibility of applying machine learning to SSI prediction in a low-resource setting and identifies areas for improvement, including larger, multi-center datasets, additional predictive variables, and standardized outcome assessment.

Chapter 6

Conclusion and Recommendation

6.1 Recommendation

Based on the findings of this study, several recommendations are proposed to improve surgical site infection (SSI) prediction and guide future research:

6.1.1 Enhance Data Collection and Feature Scope

The results of this study highlight that SSI is not strongly predictable using the variables currently available in routine hospital records. This limitation underscores the need to enrich data collection practices. Future efforts should prioritise the systematic capture of additional clinically relevant risk factors, including intraoperative variables (e.g. wound classification and contamination levels), detailed antibiotic prophylaxis timing, patient comorbidities, nutritional status, and postoperative wound assessments.

The integration of longitudinal or time-series data, including postoperative vital sign trajectories, laboratory trends, and nursing notes, would enable models to capture dynamic physiological changes that precede the onset of infection. Strengthening the granularity, completeness, and standardisation of clinical data is therefore essential for developing predictive tools with greater discriminatory power.

6.1.2 Refine Modelling Approaches for Imbalanced Data

Given the extreme class imbalance observed in this dataset, future studies should explore more advanced modelling strategies that explicitly address this challenge. Strategies specifically optimised for minority-class detection may improve sensitivity without excessively inflating false positives.

Hybrid resampling techniques, combining SMOTENC with under-sampling or synthetic data refinement, may also yield more stable minority-class representations. These refinements may enhance model robustness in rare-event prediction

6.2 Future Work

Future research should investigate hybrid modelling approaches that combine structured clinical data with unstructured sources, such as operative notes, microbiology reports, imaging findings, and nursing narratives.

Natural language processing (NLP) may uncover additional signals not captured in coded data fields. Advances in deep learning and multimodal modelling also offer opportunities to integrate text, images, and numerical data to enhance predictive accuracy. Such multimodal approaches may better reflect the multifactorial nature of SSI development.

6.3 Conclusion

The collective findings of this study indicate that SSI is not strongly predictable using the routinely collected variables available in this dataset. Across all models, discriminatory performance remained modest, precision was consistently low, and threshold adjustments did not meaningfully improve outcomes.

These results suggest that the predictors captured in routine hospital records provide only limited insight into the complex, multifactorial pathways leading to SSI. Consequently, effective early prediction may require the integration of additional clinical, microbiological, intraoperative, and environmental variables that were not available in the current dataset.

Bibliography

- [1] J. Badia, A. Casey, N. Petrosillo, P. Hudson, S. Mitchell, and C. Crosby, “Impact of surgical site infection on healthcare costs and patient outcomes: a systematic review in six european countries,” *Journal of Hospital Infection*, vol. 96, no. 1, pp. 1–15, 2017.
- [2] K. Chu, R. Maine, and M. Trelles, “Cesarean section surgical site infections in sub-saharan africa: a multi-country study from medecins sans frontieres,” *World journal of surgery*, vol. 39, no. 2, pp. 350–355, 2015.
- [3] S. Suarez-Easton, N. Zafran, G. Garmi, and R. Salim, “Postcesarean wound infection: prevalence, impact, prevention, and management challenges,” *International Journal of Women’s Health*, vol. 9, pp. 81–88, 2017.
- [4] A. Bianco, S. Roccia, C. G. A. Nobile, C. Pileggi, and M. Pavia, “Postdischarge surveillance following delivery: The incidence of infections and associated factors,” *American Journal of Infection Control*, vol. 41, no. 6, pp. 549–553, 2013.
- [5] R. Wood, A. Borodova, S. Wolter, M. N’Guessan, A. Diallo, M. Diallo, K. Heldt, C. Rocha, I. Nabé, B. Diané, M. Cherif, and S. Müller, “Surgical site infections after caesarean section across sub-sahara africa: a scoping review of prevalence and associated factors,” *Frontiers in Global Women’s Health*, vol. 6, p. 1605049, 2025.
- [6] R. Haley, D. Culver, J. White, W. Morgan, T. Emori, V. Munn, and T. Hooton, “The efficacy of infection surveillance and control programs in preventing nosocomial infections in us hospitals,” *American Journal of Epidemiology*, vol. 121, no. 2, pp. 182–205, 1985.
- [7] M. Budianu, A. Crişan, and S. Voidăzan, “Maternal sepsis—challenges in diagnosis and management—a mini-summary of the literature,” *Acta Marisiensis - Seria Medica*, vol. 70, no. 1, pp. 3–7, 2024.
- [8] A. Otu, E. Nsutebu, J. Hirst, K. Thompson, K. Walker, and S. Yaya, “How to close the maternal and neonatal sepsis gap in sub-saharan africa,” *BMJ Global Health*, vol. 5, p. e002348, 2020.
- [9] K. Islam, J. Prithula, J. Kumar, T. Tan, M. Reaz, M. Sumon, and M. Chowdhury, “Machine learning-based early prediction of sepsis using electronic health records: a systematic review,” *Journal of Clinical Medicine*, vol. 12, no. 17, p. 5658, 2023.
- [10] D. Mwandah, T. Yadesa, I. Ibanda, A. Komakech, D. Kyambadde, and J. Ngonzi, “Prevalence and factors associated with surgical site infections among mothers after cesarean section at mbarara regional referral hospital, uganda: an observational retrospective study,” *Therapeutic Advances in Infectious Disease*, vol. 11, p. 20499361241286838, 2024.
- [11] E. Smyth and A. Emmerson, “Surgical site infection surveillance,” *Journal of Hospital Infection*, vol. 45, no. 3, pp. 173–184, 2000.

- [12] T. Boerma, C. Ronsmans, D. Y. Melesse, A. J. Barros, F. C. Barros, L. Juan, A.-B. Moller, L. Say, A. R. Hosseinpour, M. Yi, *et al.*, “Global epidemiology of use of and disparities in caesarean sections,” *The Lancet*, vol. 392, no. 10155, pp. 1341–1348, 2018.
- [13] S. Mehtar, A. Wanyoro, F. Ogunsola, E. A. Ameh, P. Nthumba, C. Kilpatrick, G. Revathi, A. Antoniadou, H. Giamarelou, A. Apisarnthanarak, J. W. Ramatowski, V. D. Rosenthal, J. Storr, T. S. Osman, and J. S. Solomkin, “Implementation of surgical site infection surveillance in low- and middle-income countries: A position statement for the international society for infectious diseases,” *International Journal of Infectious Diseases*, vol. 100, pp. 123–131, 2020.
- [14] W. G. for the Clean Cut Investigators Group, M. R. Nofal, A. Tesfaye, N. Gebeyehu, N. Starr, S. Arimino, D. Chaula, C. Harrell-Shreckengost, T. Utam, R. Ambulkar, *et al.*, “Global scaling of a surgical infection prevention program to 5 low-resource countries,” *JAMA surgery*, vol. 160, no. 9, pp. 1004–1011, 2025.
- [15] S. Lakoh, L. Yi, S. Sevalie, X. Guo, O. Adekanmbi, I. O. Smalle, N. Williams, U. Barrie, C. Koroma, Y. Zhao, *et al.*, “Incidence and risk factors of surgical site infections and related antibiotic resistance in freetown, sierra leone: a prospective cohort study,” *Antimicrobial Resistance & Infection Control*, vol. 11, no. 1, p. 39, 2022.
- [16] E. Gentilotti, P. De Nardo, B. Nguhuni, A. Piscini, C. Damian, F. Vairo, Z. Chaula, P. Mencarini, P. Torokaa, A. Zumla, *et al.*, “Implementing a combined infection prevention and control with antimicrobial stewardship joint program to prevent caesarean section surgical site infections and antimicrobial resistance: a tanzanian tertiary hospital experience,” *Antimicrobial Resistance & Infection Control*, vol. 9, no. 1, p. 69, 2020.
- [17] M. Monahan, S. Jowett, T. Pinkney, P. Brocklehurst, D. G. Morton, Z. Abdali, and T. E. Roberts, “Surgical site infection and costs in low-and middle-income countries: A systematic review of the economic burden,” *PloS one*, vol. 15, no. 6, p. e0232960, 2020.
- [18] S. Zhu, X. Yang, P. Zhang, and C. Chen, “Operating room-related risk factors for surgical site infection in general surgeries: A retrospective cohort study,” *Medicine*, vol. 104, no. 37, p. e44565, 2025.
- [19] O. A. Marzoug, A. Anees, and E. M. Malik, “Assessment of risk factors associated with surgical site infection following abdominal surgery: a systematic review,” *BMJ Surgery, Interventions, & Health Technologies*, vol. 5, no. 1, p. e000182, 2023.
- [20] H. A. Megahed, A. Taki-Eldin, A. M. Abdel-monem, M. K. E.-S. Ibrahim, A. A. A. Aziz, H. A. Megahed, M. Z. Z. Farg, A. M. Kandel, A. Alemam, K. M. Salamh, *et al.*, “Risk factors for surgical site infections in elective and emergency surgeries: a prospective cohort study,” *Medical Journal of the Islamic Republic of Iran*, vol. 39, p. 55, 2025.
- [21] C. Hinson, C. Kilpatrick, K. Rasa, J. Ren, P. Nthumba, R. Sawyer, and E. Ameh, “Global surgery is stronger when infection prevention and control is incorporated: a commentary and review of the surgical infection landscape,” *BMC surgery*, vol. 24, no. 1, p. 397, 2024.
- [22] J. S. Harrop, J. C. Styliaras, Y. C. Ooi, K. E. Radcliff, A. R. Vaccaro, and C. Wu, “Contributing factors to surgical site infections,” *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, vol. 20, no. 2, pp. 94–101, 2012.
- [23] H. Rafique, “RISK FACTORS FOR SURGICAL SITE INFECTION FOLLOWING CE-SAREAN SECTION,” *BIOLOGICAL AND CLINICAL SCIENCES RESEARCH JOURNAL*, vol. 2024, no. 1, p. 1309, 2024. Publisher: Medeye Publishers.

- [24] E. Mujagic, W. R. Marti, M. Coslovsky, S. D. Soysal, R. Mechera, M. von Strauss, J. Zeindler, F. Saxer, A. Mueller, C. A. Fux, *et al.*, “Associations of hospital length of stay with surgical site infections,” *World journal of surgery*, vol. 42, no. 12, pp. 3888–3896, 2018.
- [25] O. Borens, E. Yusuf, and A. Trampuz, “Surgical site infections (ssis): risk factors and prevention strategies,” in *European Instructional Lectures: Volume 13, 2013, 14th EFORT Congress, Istanbul, Turkey*, pp. 15–24, Springer, 2013.
- [26] R. R. Fletcher, O. Olubeko, H. Sonthalia, F. Kateera, T. Nkurunziza, J. L. Ashby, R. Rivello, and B. Hedt-Gauthier, “Application of machine learning to prediction of surgical site infection,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2234–2237, IEEE, 2019.
- [27] A. C. Kiser, J. Shi, and B. T. Bucher, “An explainable long short-term memory network for surgical site infection identification,” *Surgery*, vol. 176, no. 1, pp. 24–31, 2024.
- [28] R. E. Al Mamlook, L. J. Wells, and R. Sawyer, “Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors,” *American journal of infection control*, vol. 51, no. 5, pp. 544–550, 2023.
- [29] N. Kaur and A. Kaur, “A comprehensive evaluation of ensemble machine learning algorithms for sepsis prediction: A comparative study,” in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, pp. 914–919, IEEE, 2024.
- [30] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, S. Scardapane, I. Spinelli, M. Mahmud, and A. Hussain, “Interpreting black-box models: a review on explainable artificial intelligence,” *Cognitive Computation*, vol. 16, no. 1, pp. 45–74, 2024.
- [31] C. T. Okolo, N. Dell, and A. Vashistha, “Making ai explainable in the global south: A systematic review,” in *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, pp. 439–452, 2022.
- [32] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney, “Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: A systematic review,” *Applied Sciences*, vol. 11, no. 11, 2021.
- [33] U. Pawar, D. O’Shea, S. Rea, and R. O’Reilly, “Incorporating explainable artificial intelligence (xai) to aid the understanding of machine learning in the healthcare domain,” in *Aics*, pp. 169–180, Dublin, Republic of Ireland, 2020.
- [34] A. Bell, I. Solano-Kamaiko, O. Nov, and J. Stoyanovich, “It’s just not that simple: an empirical study of the accuracy-explainability trade-off in machine learning for public policy,” in *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, pp. 248–266, 2022.
- [35] J. Wanner, L.-V. Herm, K. Heinrich, and C. Janiesch, “Stop ordering machine learning algorithms by their explainability! an empirical investigation of the tradeoff between performance and explainability,” in *Conference on e-Business, e-Services and e-Society*, pp. 245–258, Springer, 2021.
- [36] F. Obster, M. I. Ciolacu, and A. Humpe, “Balancing predictive performance and interpretability in machine learning: A scoring system and an empirical study in traffic prediction,” *IEEE Access*, vol. 12, pp. 195613–195628, 2024.

- [37] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, *et al.*, “Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai,” *Information fusion*, vol. 58, pp. 82–115, 2020.
- [38] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, “Explainable ai: A review of machine learning interpretability methods,” *Entropy*, vol. 23, no. 1, 2021.
- [39] M. Flora, C. Potvin, A. McGovern, and S. Handler, “Comparing explanation methods for traditional machine learning models part 1: an overview of current methods and quantifying their disagreement,” *arXiv preprint arXiv:2211.08943*, 2022.
- [40] H. Kaur, *Where are the Humans in Human-AI Interaction: The Missing Human-Centered Perspective on Interpretability Tools for Machine Learning*. PhD thesis, 2023.
- [41] A. F. Markus, J. A. Kors, and P. R. Rijnbeek, “The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies,” *Journal of Biomedical Informatics*, vol. 113, p. 103655, 2021.
- [42] R. Zhou and T. Hu, “Evolutionary approaches to explainable machine learning,” in *Handbook of evolutionary machine learning*, pp. 487–506, Springer, 2023.
- [43] A. S. Shaikh, R. M. Samant, K. S. Patil, N. R. Patil, and A. R. Mirkale, “Review on explainable ai by using lime and shap models for healthcare domain,” *Int. J. Comput. Appl*, vol. 185, pp. 18–23, 2023.
- [44] O. A. Akande, “Leveraging explainable ai models to improve predictive accuracy and ethical accountability in healthcare diagnostic decision support systems,” *World Journal of Advanced Research and Reviews*, vol. 8, no. 2, pp. 415–434, 2020.
- [45] B. Allegranzi, B. Zayed, P. Bischoff, N. Z. Kubilay, S. de Jonge, F. de Vries, S. M. Gomes, S. Gans, E. D. Wallert, X. Wu, *et al.*, “New who recommendations on intraoperative and postoperative measures for surgical site infection prevention: an evidence-based global perspective,” *The Lancet Infectious Diseases*, vol. 16, no. 12, pp. e288–e303, 2016.
- [46] J. X. C. Ke, A. DhakshinaMurthy, R. B. George, and P. Branco, “The effect of resampling techniques on the performances of machine learning clinical risk prediction models in the setting of severe class imbalance: development and internal validation in a retrospective cohort,” *Discover Artificial Intelligence*, vol. 4, no. 1, p. 91, 2024.
- [47] S. M. Simon, P. Glaum, and F. S. Valdovinos, “Interpreting random forest analysis of ecological models to move from prediction to explanation,” *Scientific reports*, vol. 13, no. 1, p. 3881, 2023.
- [48] D. Lee, L. Kwan, E. Nguyen, I. Pluym, and M. Tandel, “Maternal morbidity rates relative to length of labor in a multiethnic patient population [id: 1337798],” *Obstetrics & Gynecology*, vol. 141, no. 5S, p. 42S, 2023.
- [49] M. Ghanem, A. K. Ghaith, V. G. El-Hajj, A. Bhandarkar, A. De Giorgio, A. Elmi-Terander, and M. Bydon, “Limitations in evaluating machine learning models for imbalanced binary outcome classification in spine surgery: a systematic review,” *Brain sciences*, vol. 13, no. 12, p. 1723, 2023.

Appendix A

Budget

	Year 1 (US \$)	Year 2 (US \$)	Year 3 (US \$)	Year 4 (US \$)	Year 5 (US \$)	TOTAL
Personnel	500	0	0	0	0	500
Travel *	300	0	0	0	0	300
Materials and Supplies	100	0	0	0	0	100
Administration	1500	0	0	0	0	1500
Results dissemination	1000	0	0	0	0	1000
Contingency	100	0	0	0	0	100
Reimbursement and Time Compensations	800	0	0	0	0	800
Other	0	0	0	0	0	0
Total	3500	0	0	0	0	4300

Figure A.1: Budget allocation for the study activities.

Appendix B

Research Project timelines

A Gantt Chart may be useful to show the phases, tasks and dependencies on one page.

Activity	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
REC approval & administrative clearance						
Data access & verification						
Data cleaning & preprocessing						
Exploratory data analysis (EDA)						
Model development						
Model evaluation						
Model interpretability analysis						
Results interpretation & writing						
Dissemination & final reporting						

Figure B.1: Activity Timeline

Appendix B

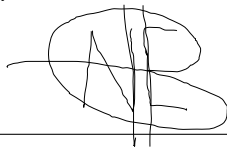
Declaration of AI use

Note: Briefly explain the ways that you have ethically used AI as your “study buddy” according to the guidelines stated in section . Explain which AI tools you have used and for what purposes. Describe what output from the tool/service and the prompts used. In case you used AI to edit, please indicate how the content was modified, and submit the original version of the work before AI editing.

Declaration Statement:

I, *Emmanuel Nkurunziza*, declare that I used

- i. *CoPilot* to *To troubleshoot code* and *Evaluate this code and identify why there is an error*. I modified the outputs in *I corrected the wrong variable name..*
- ii. *(chat GPT* to *(trouble shoot and align the images in this work into Latex)* and *(Why are the images in latex not appearing under the subsection headers?)*. I modified the outputs in Latex *(by adding a Float barrier command as guided by chat GPT)*.



Signature: _____ Date: 03/04/2026

Appendix D

List of 10 Major Journal Publications reviewed

1. Badia, J., et al., Impact of surgical site infection on healthcare costs and patient outcomes: a systematic review in six European countries. *Journal of Hospital Infection*, 2017. 96(1): p. 1-15.
2. Chu, K., R. Maine, and M. Trelles, Cesarean section surgical site infections in sub-Saharan Africa: a multi-country study from Medecins Sans Frontieres. *World journal of surgery*, 2015. 39(2): p. 350-355.
3. Zuarez-Easton, S., et al., Postcesarean wound infection: prevalence, impact, prevention, and management challenges. *International journal of women's health*, 2017: p. 81-88.
4. Wood, R., et al., Surgical site infections after caesarean section across sub-Sahara Africa: a scoping review of prevalence and associated factors. *Frontiers in Global Women's Health*, 2025. 6: p. 1605049.
5. Haley, R.W., et al., The efficacy of infection surveillance and control programs in preventing nosocomial infections in US hospitals. *American journal of epidemiology*, 1985. 121(2): p. 182-205.
6. Budianu, M.A., A.I. Crişan, and S. Voidăzan, Maternal sepsis-challenges in diagnosis and management-a mini-summary of the literature. *Acta Marisiensis-Seria Medica*, 2024. 70(1).
7. Otu, A., et al., How to close the maternal and neonatal sepsis gap in sub-Saharan Africa. 2020, BMJ Publishing Group Ltd.
8. Islam, K.R., et al., Machine learning-based early prediction of sepsis using electronic health records: a systematic review. *Journal of clinical medicine*, 2023. 12(17): p. 5658.
9. Mwandah, D.C., et al., Prevalence and factors associated with surgical site infections among mothers after cesarean section at Mbarara Regional Referral Hospital, Uganda: an observational retrospective study. *Therapeutic Advances in Infectious Disease*, 2024. 11: p. 20499361241286838.
10. Smyth, E. and A. Emmerson, Surgical site infection surveillance. *Journal of Hospital Infection*, 2000. 45(3): p. 173-184.

Appendix G

Research poster layout

– Please include a one-page-poster layout summarizing your research concept, the problem being addressed, the objectives, the alignment with national and global agenda, the proposed methodology and expected results. you may include two major references used as a basis for your research.

Research Poster

AUTHORS & INSTITUTION

Emmanuel Nkurunziza

Uganda Christian University

Interpretable Predictive Modeling of Postpartum Surgical Site Infections

ABSTRACT

This study applies interpretable ML to predict postpartum surgical site infections (SSIs) using routine clinical data from Ugandan hospitals.

Findings show modest predictive performance but highlight key risk factors and demonstrate the value of XAI for early detection and improved maternal care.

INTRODUCTION

Postpartum surgical site infections often occur after discharge, where follow-up systems are weak and detection is delayed and symptom-driven. This study explores the use of interpretable machine learning to enable early identification of high-risk patients, supporting timely intervention and improved maternal outcomes.

METHODOLOGY

8,918 caesarean deliveries from 3 Ugandan regional referral hospitals

Preprocessing: Cleaning, missing value handling, feature engineering, encoding
Class imbalance: Addressed using SMOTENC

Models: Logistic Regression, Random Forest, SVM (RBF), XGBoost

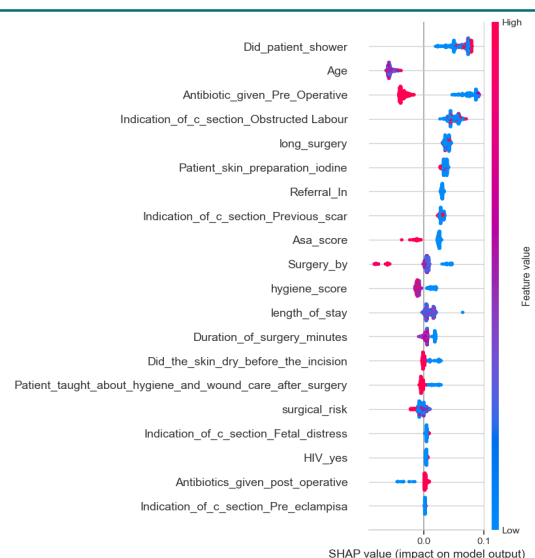
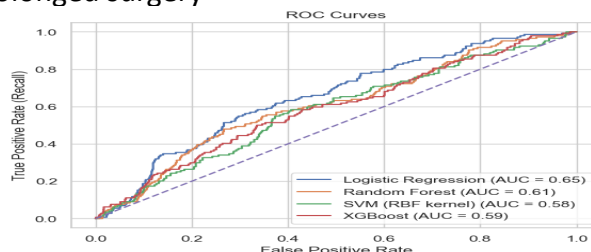
Evaluation: Recall, F1-score, ROC-AUC, Precision-Recall

DISCUSSION

The models identified clinically meaningful risk factors such as prolonged surgery, obstructed labour, and emergency procedures, consistent with existing literature. Predictive performance was modest due to limited data quality, missing key variables, and severe class imbalance, with LR achieving high recall but low precision. Despite these limitations, explainability (SHAP) enhanced transparency and trust by revealing how predictions were made.

RESULTS

Model performance was modest due to limited predictive signal in routine clinical data. Logistic Regression achieved very high recall (0.9931; ROC-AUC = 0.6528), while XGBoost showed higher accuracy (0.7932) but low recall (0.2639). SHAP identified key risk factors including poor preoperative hygiene, obstructed labour, and prolonged surgery



CONCLUSION

Explainable ML models can help identify early SSI risk in low-resource settings but their performance is limited by the sparse routine clinical data available. LR stands out as a useful high-sensitivity screening tool for early detection. Better results will require richer clinical datasets to enable practical use in maternal health surveillance.